

On MMSE Properties and I-MMSE

Implications in Parallel MIMO Gaussian Channels

Ronit Bustin, Miquel Payaró, Daniel P. Palomar, and Shlomo Shamai (Shitz)

Abstract

The scalar additive Gaussian noise channel has the “single crossing point” property between the minimum-mean square error (MMSE) in the estimation of the input given the channel output, assuming a Gaussian input to the channel, and the MMSE assuming an arbitrary input. This paper extends the result to the parallel MIMO additive Gaussian channel in three phases: i) The channel matrix is the identity matrix, and we limit the Gaussian input to a vector of Gaussian i.i.d. elements. The “single crossing point” property is with respect to the snr (as in the scalar case). ii) The channel matrix is arbitrary, the Gaussian input is limited to an independent Gaussian input. A “single crossing point” property is derived for each diagonal element of the MMSE matrix. iii) The Gaussian input is allowed to be an arbitrary Gaussian random vector. A “single crossing point” property is derived for each eigenvalue of the MMSE matrix.

These three extensions are then translated to new information theoretic properties on the mutual information, using the fundamental relationship between estimation theory and information theory. The results of the last phase are also translated to a new property of Fisher’s information. Finally, the applicability of all three extensions on information theoretic problems is demonstrated through: a proof of a special case of Shannon’s vector EPI, a converse proof of the capacity region of the parallel *degraded* MIMO broadcast channel (BC) under per-antenna power constraints and under covariance constraints, and a converse proof of the capacity region of the compound parallel *degraded* MIMO BC under covariance constraint.

I. INTRODUCTION

This paper considers parallel multiple-input multiple-output (MIMO) channels, with an arbitrary input distribution and additive standard Gaussian noise. These channels are a subset of the important family of MIMO additive Gaussian noise channels, which have been extensively investigated in the literature. For most Gaussian channel models studied in information theory, Gaussian signaling happens to be optimal, from point-to-point channels, to multiple-access channels (MAC), and broadcast channels (BC) [1, Ch. 9 and 15] [2]. The methods used to prove this optimality were not easy to come across, even when considering scalar Gaussian channels. For example, in order to prove that Gaussian inputs are optimal for the scalar Gaussian BC, Bergmans employed Shannon’s entropy power inequality (EPI) [3]. The solution for the MIMO Gaussian BC came only 30 years later in [2], using a new enhancement approach. Since then, several other proofs were derived, using different tools, such as, the extremal inequality in [4], the de Bruijn identity in coordination with Dembo’s inequality in [5], and the “single crossing

point” property presented by Guo et al. in [6]. The “single crossing point” stemmed from the I-MMSE relationship, a fundamental relationship between estimation theory and information theory revealed by Guo, Shamai, and Verdú in [7].

The relationship between estimation theory and information theory goes back to the late 1950’s, when Stam [8] used the de Bruijn’s identity to prove Shannon’s EPI, and then in the early 1970’s when the mutual information was represented as a function of the causal filtering error by Duncan [9] and Kadota, Zakai and Ziv [10]. The I-MMSE relationship, given for discrete-time and continuous-time, scalar and vector additive Gaussian noise channels, deepens the connection between these two fields. Specifically, for a scalar additive Gaussian noise channel,

$$Y = \sqrt{\text{snr}}X + N \quad (1)$$

where N is standard Gaussian additive noise, then, regardless of the input distribution of X , the mutual information, $I(X; Y)$, and minimum-mean-square error (MMSE) in the estimation of X given the observation Y , $\text{mmse}(X, \text{snr})$, are related (assuming real-valued inputs/outputs) by

$$D_{\text{snr}} I(X; Y) = \frac{1}{2} \text{mmse}(X, \text{snr}) \quad (2)$$

where D_{snr} is the derivative with respect to snr , and

$$\text{mmse}(X, \text{snr}) = \mathbb{E}\{(X - \mathbb{E}\{X \mid \sqrt{\text{snr}}X + N\})^2\}. \quad (3)$$

The work in [7] has been extended in several directions, among which we have: the additive Poisson noise channel [11], [12], the general additive noise channel [13], arbitrary channels [14], representation of the relative entropy as a function of the difference between the mismatched MMSE and the matched MMSE in [15], [16], and others. One important extension, on which we heavily rely, is the one done by Palomar and Verdú in [17], where they obtain the gradient of the mutual information with respect to different parameters of the MIMO channel.

Going back to the “single crossing point” property, one of the goals in [6] was to show the applicability of the I-MMSE relationship as a tool to solve information-theoretic problems. Specifically, the authors of [6] examined the scalar Gaussian BC and gave an alternative proof for the optimality of Gaussian inputs. In order to show this, Guo et al. defined the following function in [6]:

$$f(X, \gamma) = (1 + \gamma)^{-1} - \text{mmse}(X, \gamma) \quad (4)$$

where the simplified notation $f(\gamma)$ will be used when there is no confusion about the distribution of X . It was shown that $f(\gamma)$ has *at most* a single crossing point of the horizontal axis. In other words, the first term, which is the MMSE assuming a standard Gaussian input, may be smaller than the second term in some range of snr values (note that the parameter γ is the snr); however, once the two terms are equal, at some γ_0 , the MMSE of the standard Gaussian input remains greater than the MMSE of the arbitrary input for all $\gamma > \gamma_0$, and the function remains nonnegative. This property together with the I-MMSE relationship, provides the missing link to derive a simple and elegant converse proof of the capacity region of the scalar Gaussian BC.

The “single crossing point” was derived only for the scalar additive Gaussian channel, as can be seen from the definition of the function $f(\gamma)$. The motivation of this work is to extend this property to the vector Gaussian channel. This extension is done in three phases. First, we consider random input vectors instead of random input scalars, but keep the dependence on a scalar quantity – the channel snr. In this setting we also limit the Gaussian input to a vector of Gaussian i.i.d. elements. In the second and third phases we consider dependence on a vector quantity – a parallel channel matrix. The difference between the second and third phases is that in the second phase we limit the Gaussian input to an independent random vector, whereas in the third phase the Gaussian input is arbitrary. In all three phases, the general channel model considered is the following:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N} \quad (5)$$

where \mathbf{N} is a standard Gaussian random vector, and \mathbf{H} is a square and diagonal channel matrix known to the receiver(s). In the vector case, the scalar MMSE does not capture all the needed information, and we need to resort to the matrix extension, the MMSE matrix defined as

$$\mathbf{E}_\mathbf{X} = \mathbb{E}\{(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{H}\mathbf{X} + \mathbf{N}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{H}\mathbf{X} + \mathbf{N}\})^\top\} \quad (6)$$

from which we can see that, in general, the MMSE matrix $\mathbf{E}_\mathbf{X}$ depends on the channel $\mathbf{E}_\mathbf{X} = \mathbf{E}_\mathbf{X}(\mathbf{H})$, but whenever the channel coefficients depend on other parameters $\mathbf{H} = \mathbf{H}(\phi)$, we will write $\mathbf{E}_\mathbf{X}(\phi)$. Observe that the standard scalar MMSE value in the vector case can be easily recovered from the MMSE matrix as follows:

$$\text{mmse}(\mathbf{X}, \text{snr}) = \mathbb{E}\{\|\mathbf{X} - \mathbb{E}\{\mathbf{X} | \sqrt{\text{snr}}\mathbf{X} + \mathbf{N}\}\|^2\} = \text{Tr}(\mathbf{E}_\mathbf{X}(\sqrt{\text{snr}}\mathbf{I}_n)). \quad (7)$$

For the important case when the input distribution of \mathbf{X} is Gaussian with covariance matrix $\mathbf{R}_{\mathbf{X}_G}$ we will use the following notation:

$$\mathbf{E}_G(\mathbf{R}_{\mathbf{X}_G}, \mathbf{H}) = (\mathbf{R}_{\mathbf{X}_G}^{-1} + \mathbf{H}^\top \mathbf{H})^{-1} \quad (8)$$

where we assumed that $\mathbf{R}_{\mathbf{X}_G}$ is of full rank. As in the case of $\mathbf{E}_\mathbf{X}$, whenever the channel coefficients depend on other parameters $\mathbf{H} = \mathbf{H}(\phi)$, we will write $\mathbf{E}_G(\mathbf{R}_{\mathbf{X}_G}, \phi)$. Another important quantity is the MMSE given for a specific output, $\mathbf{Y} = \mathbf{y}$, defined as:

$$\Phi_\mathbf{X}(\mathbf{y}) = \mathbb{E}\{(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{y}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{y}\})^\top | \mathbf{y}\}. \quad (9)$$

Although not specified explicitly, $\Phi_\mathbf{X}(\mathbf{y})$ depends on the channel matrix/parameters. Note that $\mathbf{E}_\mathbf{X}(\mathbf{H}) = \mathbb{E}\{\Phi_\mathbf{X}(\mathbf{Y})\}$. Interestingly, when the input distribution of \mathbf{X} is Gaussian, $\Phi_\mathbf{X}(\mathbf{y})$ is independent of \mathbf{y} and the following equality holds for all \mathbf{y} : $\mathbf{E}_G(\mathbf{R}_{\mathbf{X}_G}, \mathbf{H}) = \Phi_\mathbf{X}(\mathbf{y})$. Finally, given all these quantities we can define the main player in this work: the MMSE matrix difference (analog to $f(\gamma)$ in the scalar case)

$$\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, \phi) = \mathbf{E}_G(\mathbf{R}_{\mathbf{X}_G}, \phi) - \mathbf{E}_\mathbf{X}(\phi) \quad (10)$$

$$= \mathbf{E}_{\mathbf{X}_G}(\phi) - \mathbf{E}_\mathbf{X}(\phi) \quad (11)$$

$$= (\mathbf{R}_{\mathbf{X}_G}^{-1} + (\mathbf{H}(\phi))^\top \mathbf{H}(\phi))^{-1} - \mathbf{E}_\mathbf{X}(\phi) \quad (12)$$

where, similarly to the scalar case in (4), we will use the simplified notation $\mathbf{Q}(\phi)$ when the distribution of \mathbf{X} and the covariance matrix of the Gaussian distribution $\mathbf{R}_{\mathbf{X}_G}$ are clear from the context. Note that there is *no* requirement that the covariance of the random vector \mathbf{X} be equal to $\mathbf{R}_{\mathbf{X}_G}$.

As it has already been pointed out, the extension from scalar-to-vector is done in three phases. In the first step the dependence remains on a scalar parameter – the snr. This is obtained by setting $\mathbf{H} = \mathbf{I}$ in the general MIMO model in (5). We further limit our observation to the comparison of an arbitrary input distribution with the subset of Gaussian random vectors with i.i.d. elements. For this case, we show that the “single crossing point” property extends smoothly to any linear combination of $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, \phi)$ with a positive semidefinite matrix. Although this is the simplest scalar-to-vector extension, the proof is not straightforward. In order to demonstrate the applicability of this result we extend the proof of a special case of Shannon’s EPI, done in [6], to the vector case.

Proceeding with the scalar-to-vector extension, we assume that the channel matrix, \mathbf{H} , is parallel, thus our dependence is now on a vector parameter. In this setting we have two distinguishable results, given in phases two and three, that cannot be trivially deduced from each other. In phase two, we limit the Gaussian distribution, to which we compare, to any independent Gaussian distribution characterized by its diagonal covariance matrix, $\mathbf{\Lambda}_{\mathbf{X}_G}$. Under this assumption we show that a “single crossing point” property exists for each and every diagonal element of $\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, \phi)$. Together with the I-MMSE relationship, this result provides some interesting properties of the mutual information, and its applicability is demonstrated by providing a simple converse proof for the parallel Gaussian BC capacity region under per-antenna power constraints.

The third phase, which is the main result of this work, does not require any further assumptions (apart from the diagonal channel matrix). That is, we compare an arbitrary input distribution with any general Gaussian input distribution, with covariance $\mathbf{R}_{\mathbf{X}_G}$. In this setting we show that a “single crossing point” property exists for each and every eigenvalue of the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, \phi)$. The applicability of this result is demonstrated with two information-theoretic problems: the converse proof of the parallel Gaussian BC capacity region under covariance constraint and the converse proof of the compound parallel Gaussian BC capacity region under covariance constraint.

Much of this work regards the behavior of functions around zeros, the existence and amount of actual crossings of the horizontal axis. Thus, before proceeding with the technical content of the paper and, in order to make these observations rigorous, we require the next definitions which will be used throughout the paper.

Definition 1: Given a function $h(t)$ continuous within the neighborhood of t_0 , we say that a negative-to-nonnegative zero crossing occurs at $t = t_0$ if, and only if, $h(t_0) = 0$ and there exists a positive value ϵ such that $h(t) < 0$ for $t \in (t_0 - \epsilon, t_0)$ and $h(t) \geq 0$ for $t \in (t_0, t_0 + \epsilon)$.

Definition 2: Given a function $h(t)$ continuous within the neighborhood of t_0 , we say that a nonnegative-to-negative zero crossing occurs at $t = t_0$ if, and only if, $h(t_0) = 0$ and there exists a positive value ϵ such that $h(t) \geq 0$ for $t \in (t_0 - \epsilon, t_0)$ and $h(t) < 0$ for $t \in (t_0, t_0 + \epsilon)$.

Similar definitions can be given for positive-to-nonpositive and nonpositive-to-positive zero crossings. Another required definition is the following:

Definition 3: Given a function $h(t)$ continuous within the neighborhood of t_0 , we say that a negative-zero-positive

crossing occurs at $t = t_0$ if, and only if, a negative-to-nonnegative zero crossing occurs as $t = t_0$ and there exists a positive δ such that $h(t) = 0$ for $t \in (t_0, t_0 + \delta)$ and a nonpositive-to-positive zero crossing occurs as $t_0 + \delta$.

Similarly we can define a positive-zero-negative crossing.

The remaining of this paper is organized as follows: Section II considers the first phase of our extension from scalar-to-vector, in which case the dependence is on the scalar parameter, snr . In Section III we provide the framework in which we handle the assumption of a parallel channel matrix, \mathbf{H} . This framework is relevant for phases two and three of our scalar-to-vector extension. In Section IV we consider phase two of our extension, where we limit our observations to an independent Gaussian input distribution. Section V considers phase three, where we compare the arbitrary input to any general Gaussian input distribution.

Notation: Straight boldface denotes multivariate quantities such as vectors (lowercase) and matrices (uppercase). Uppercase italics denotes random variables (boldface if we consider random vectors rather than random variables), and their realizations are represented by lowercase italics. The set of n -dimensional positive semidefinite matrices is denoted by \mathbb{S}_+^n . The elements of a matrix \mathbf{A} are represented by $[\mathbf{A}]_{ij}$. The operator $\text{diag}(\mathbf{A})$ represents a column vector with the diagonal entries of matrix \mathbf{A} , and $\text{Diag}(\mathbf{a})$ represents a diagonal matrix whose non-zero elements are given by the elements of vector \mathbf{a} . The superscript $(\cdot)^T$ denotes the transpose. The operator $\text{Tr}(\cdot)$ denotes the trace function, and $|\cdot|$ denotes the determinant function. The operator $D_\gamma \mathbf{A}$ denotes the Jacobian matrix of \mathbf{A} with respect to γ [18].

Note that we also consider the conditioned version of the above defined quantities. That is, when the random vector \mathbf{X} depends on the random vector \mathbf{U} , we require, for example, a conditioned version for the MMSE and the matrix \mathbf{Q} given for a specific value of $\mathbf{U} = \mathbf{u}$. In this case both quantities depend on an additional parameter \mathbf{u} , i.e., $\mathbf{E}_{\mathbf{X}|\mathbf{U}}(\phi, \mathbf{u})$ and $\mathbf{Q}(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{R}_{\mathbf{X}_G}, \phi)$ (the precise definitions given in Section IV-B).

II. THE SCALAR MIMO CHANNEL

As pointed out in the introduction, we begin our study with the simplest multivariate extension of the result in [6, Prp. 16], that is, we consider that the scalar random variables involved in the model in (1) become random vectors. In other words, in this section, we consider the following model:

$$\mathbf{Y} = \sqrt{\text{snr}}\mathbf{X} + \mathbf{N} \quad (13)$$

where the input random vector $\mathbf{X} \in \mathbb{R}^n$ is arbitrarily distributed and $\mathbf{N} \in \mathbb{R}^n$ follows a standard Gaussian distribution. Observe that (13) is obtained by setting $\mathbf{H} = \sqrt{\text{snr}}\mathbf{I}_n$ in the vector model in (5).

Moreover, we further limit our discussion in this section to the comparison with a Gaussian input with i.i.d. elements, i.e., we assume that $\mathbf{R}_{\mathbf{X}_G} = \sigma^2\mathbf{I}_n$.

Thus, for the settings in this section, the general MMSE matrix difference function in (12) simplifies to

$$\mathbf{Q}(\mathbf{X}, \sigma^2\mathbf{I}_n, \gamma) = \frac{\sigma^2}{1 + \sigma^2\gamma}\mathbf{I}_n - \mathbf{E}_{\mathbf{X}}(\gamma) \quad (14)$$

where γ plays the role of the estimation snr .

A. A Single Crossing Point

Motivated by the “single crossing property” of $f(X, \gamma)$ presented in [6, Prp. 16], an immediate question that comes to mind is “does this property extend to the MIMO scenario?” Our hypothesis was that for the setting in (13) this property will have a simple extension. Thus, we examine the simplest scalar function of the MMSE matrix difference function of (14), that is, we consider some linear combination of it. Accordingly, we define

$$q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma) = \text{Tr}(\mathbf{A}\mathbf{Q}(\mathbf{X}, \sigma^2\mathbf{I}_n, \gamma)) \quad (15)$$

$$= \frac{\sigma^2}{1 + \sigma^2\gamma} \text{Tr}(\mathbf{A}) - \text{Tr}(\mathbf{A}\mathbf{E}_{\mathbf{X}}(\gamma)) \quad (16)$$

where \mathbf{A} is a weighting matrix.

The “single crossing point” property of $f(\gamma)$ extends naturally to the function $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$, for a specific subset of matrices \mathbf{A} . This result is given in the next theorem.

Theorem 1: Let $\mathbf{A} \in \mathbb{S}_+^n$ be a positive semidefinite matrix. Then, the function $\gamma \mapsto q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$, defined in (16), has no nonnegative-to-negative zero crossings and, at most, a single negative-to-nonnegative zero crossing in the range $\gamma \in [0, \infty)$.

Moreover, assume $\text{snr}_0 \in [0, \infty)$ is a negative-to-nonnegative crossing point. Then,

- 1) $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, 0) \leq 0$.
- 2) $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$ is a strictly increasing function in the range $\gamma \in [0, \text{snr}_0)$.
- 3) $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma) \geq 0$ for all $\gamma \in [\text{snr}_0, \infty)$.
- 4) $\lim_{\gamma \rightarrow \infty} q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma) = 0$.

Proof: We start with the following three lemmas that are instrumental for this proof.

Lemma 1: Let $\mathbf{A} \in \mathbb{S}_+^n$ be a positive semidefinite matrix and let the random vector $\mathbf{X} \in \mathbb{R}^n$ be arbitrarily distributed. Then, we can always find a random vector $\hat{\mathbf{X}} \in \mathbb{R}^n$ such that the number of nonnegative-to-negative and negative-to-nonnegative zero crossings of $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$ is the same as those of $q_{\mathbf{I}_n}(\hat{\mathbf{X}}, \sigma^2, \gamma)$.

Proof: See Appendix A1. ■

Lemma 2: Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector such that $\text{Tr}(\mathbf{R}_{\mathbf{X}})/n \leq \sigma^2$. Then, for every $\gamma \geq 0$, we have

$$\frac{\text{Tr}(\mathbf{E}_{\mathbf{X}}(\gamma))}{n} \leq \frac{\sigma^2}{1 + \sigma^2\gamma} \quad (17)$$

with equality if and only if \mathbf{X} is a Gaussian vector with i.i.d. elements of variance σ^2 .

Proof: See Appendix A2. ■

Lemma 3: Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a square matrix. The derivative of the function $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$ with respect to γ is given by

$$D_{\gamma}q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma) = \text{Tr}(\mathbf{A}\mathbf{E}\{\Phi_{\mathbf{X}}(\mathbf{Y})^2\}) - \frac{\sigma^4}{(1 + \sigma^2\gamma)^2} \text{Tr}(\mathbf{A}). \quad (18)$$

Proof: See Appendix A3. ■

With these three lemmas at hand, we are now ready to continue with the proof of Theorem 1.

Since we are assuming that the matrix \mathbf{A} is positive semidefinite and the distribution of \mathbf{X} is arbitrary, from Lemma 1, we see that we can restrict our study of $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$ to that of $q_{\mathbf{I}_n}(\hat{\mathbf{X}}, \sigma^2, \gamma)$. For the sake of simplicity, throughout this proof we will use $q(\sigma^2, \gamma) = q_{\mathbf{I}_n}(\hat{\mathbf{X}}, \sigma^2, \gamma)$.

Now, according to Lemma 2, for the case where $\text{Tr}(\mathbf{R}_{\mathbf{X}})/n < \sigma^2$, the function $q(\sigma^2, \gamma)$ has no zeros and the statement in Theorem 1 is true. In addition, if \mathbf{X} is Gaussian distributed with covariance matrix equal to $\sigma^2 \mathbf{I}_n$, then $q(\sigma^2, \gamma) = 0, \forall \gamma$, which also fulfills Theorem 1.

Thus, from this point, we can assume that $\text{Tr}(\mathbf{R}_{\mathbf{X}})/n \geq \sigma^2$ and that \mathbf{X} is not a Gaussian vector with covariance matrix $\sigma^2 \mathbf{I}_n$. Now, for $\gamma = 0$ we have $q(\sigma^2, 0) = \sigma^2 - \text{Tr}(\mathbf{R}_{\mathbf{X}})/n \leq 0$ as required.

From the smoothness of $q(\sigma^2, \gamma)$ as a function of γ , and done in [6, Prp. 16], in order to prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $q(\sigma^2, \gamma)$ can occur, we only need to show that the derivative of $q(\sigma^2, \gamma)$ is positive for all values of γ for which $q(\sigma^2, \gamma) < 0$. Observe that $q(\sigma^2, \gamma) < 0$ implies that

$$n \frac{\sigma^2}{1 + \sigma^2 \gamma} < \text{Tr}(\mathbf{E}_{\mathbf{X}}(\gamma)) = \text{Tr}(\mathbf{E}\{\Phi_{\mathbf{X}}(\mathbf{Y})\}). \quad (19)$$

Now, particularizing Lemma 3 for $\mathbf{A} = \mathbf{I}_n$, we have that

$$D_{\gamma} q(\sigma^2, \gamma) = \text{Tr}(\mathbf{E}\{\Phi_{\mathbf{X}}(\mathbf{Y})^2\}) - n \frac{\sigma^4}{(1 + \sigma^2 \gamma)^2} \quad (20)$$

$$> \text{Tr}(\mathbf{E}\{\Phi_{\mathbf{X}}(\mathbf{Y})^2\}) - \frac{(\text{Tr}(\mathbf{E}\{\Phi_{\mathbf{X}}(\mathbf{Y})\}))^2}{n} \quad (21)$$

$$= \mathbf{1}^T \mathbf{E}\{\Phi_{\mathbf{X}}(\mathbf{Y}) \circ \Phi_{\mathbf{X}}(\mathbf{Y})\} \mathbf{1} - \mathbf{1}^T \frac{\mathbf{E}\{\text{diag}(\Phi_{\mathbf{X}}(\mathbf{Y}))\} \mathbf{E}\{\text{diag}(\Phi_{\mathbf{X}}(\mathbf{Y}))^T\} \mathbf{1}}{n} \quad (22)$$

$$\geq \mathbf{E} \left\{ \mathbf{1}^T \left(\Phi_{\mathbf{X}}(\mathbf{Y}) \circ \Phi_{\mathbf{X}}(\mathbf{Y}) - \frac{\text{diag}(\Phi_{\mathbf{X}}(\mathbf{Y})) \text{diag}(\Phi_{\mathbf{X}}(\mathbf{Y}))^T}{n} \right) \mathbf{1} \right\} \quad (23)$$

$$\geq 0 \quad (24)$$

where (21) follows directly from (19); in (22) we have defined $\mathbf{1}$ as the column vector whose entries are all ones, and we used \circ to denote the Schur product; and (23) follows from Jensen's inequality; finally, (24) follows from [14, Prp. H.9].

Observe that the inequality in (24), which holds for values of γ such that $q(\sigma^2, \gamma) < 0$, also proves the second item in Theorem 1 and the third one follows directly from the inexistence of nonnegative-to-negative zero crossings. Furthermore, regarding the fourth item, it is clear that $\lim_{\gamma \rightarrow \infty} q(\sigma^2, \gamma) = 0$, as both terms in $q(\sigma^2, \gamma)$ tend to zero. ■

Remark 1: Note that the above theorem also holds for the normalized function, $\frac{1}{n} q(\sigma^2, \gamma)$. Specifically, for the case of $\mathbf{A} = \mathbf{I}_n$, this is simply the difference between the MMSE of a general Gaussian random variable, with variance σ^2 , and the *average* MMSE of the n elements of the random vector \mathbf{X} .

Remark 2: For negative semidefinite \mathbf{A} it can easily be seen from the proof of Lemma 1 that $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$ has the inverse properties, since it is a mirroring of some $q_{\mathbf{I}_n}(\hat{\mathbf{X}}, \sigma^2, \gamma)$ over the x-axis. This is to say, that it has at

most a single positive-to-nonpositive zero crossing and, if such crossing exists, $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$ will be nonnegative at $\gamma = 0$, strictly decreasing up to the crossing, nonpositive after the crossing, and will tend to zero as $\gamma \rightarrow \infty$.

Remark 3: For indefinite \mathbf{A} , “single crossing point” properties, such as those shown in Theorem 1, do not hold in general.

B. Application: A Proof of a Special Case of Shannon’s Vector EPI

We now show that Theorem 1 can be used to prove a special case of Shannon’s EPI [1, Th. 17.7.3], similarly as it was done in [6] for the scalar case. Precisely, we will show that

$$\exp\left(\frac{2}{n}h(\mathbf{X} + \mathbf{N})\right) \geq \exp\left(\frac{2}{n}h(\mathbf{X})\right) + 2\pi e|\mathbf{R}_{\mathbf{N}}|^{\frac{1}{n}} \quad (25)$$

for any independent n -dimensional vectors \mathbf{X} and \mathbf{N} as long as the differential entropy of \mathbf{X} is well-defined and \mathbf{N} is Gaussian distributed with a positive definite covariance matrix $\mathbf{R}_{\mathbf{N}}$.

We define \mathbf{Z} to be an n -dimensional Gaussian vector with covariance $\mathbf{R}_{\mathbf{Z}} = \mathbf{R}_{\mathbf{N}}$ and independent of both \mathbf{X} and \mathbf{N} . Thus, without making any assumptions on the covariance matrix of \mathbf{X} , we can find an $\alpha \in [0, \infty)$ such that the following equality holds:

$$h(\mathbf{X}) = h(\alpha\mathbf{Z}) = \frac{1}{2}\log((2\pi e)^n \alpha^{2n} |\mathbf{R}_{\mathbf{N}}|). \quad (26)$$

Since $\mathbf{R}_{\mathbf{N}}$ is positive definite there exists an invertible matrix \mathbf{V} such that $\mathbf{R}_{\mathbf{N}} = \mathbf{V}\mathbf{V}^T$. Defining $\tilde{\mathbf{X}} = \mathbf{V}^{-1}\mathbf{X}$, $\tilde{\mathbf{Z}} = \mathbf{V}^{-1}\mathbf{Z}$ and $\tilde{\mathbf{N}} = \mathbf{V}^{-1}\mathbf{N}$ we have the following chain of equalities:

$$\Delta I(\text{snr}) = I(\alpha\mathbf{Z}; \sqrt{\text{snr}}\alpha\mathbf{Z} + \mathbf{N}) - I(\mathbf{X}; \sqrt{\text{snr}}\mathbf{X} + \mathbf{N}) \quad (27)$$

$$= I(\alpha\tilde{\mathbf{Z}}; \sqrt{\text{snr}}\alpha\tilde{\mathbf{Z}} + \tilde{\mathbf{N}}) - I(\tilde{\mathbf{X}}; \sqrt{\text{snr}}\tilde{\mathbf{X}} + \tilde{\mathbf{N}}) \quad (28)$$

$$= h(\sqrt{\text{snr}}\alpha\tilde{\mathbf{Z}} + \tilde{\mathbf{N}}) - h(\sqrt{\text{snr}}\tilde{\mathbf{X}} + \tilde{\mathbf{N}}) \quad (29)$$

$$= \frac{1}{2} \int_0^{\text{snr}} (\text{mmse}(\alpha\tilde{\mathbf{Z}}, \gamma) - \text{mmse}(\tilde{\mathbf{X}}, \gamma)) d\gamma \quad (30)$$

$$= \frac{1}{2} \int_0^{\text{snr}} \text{Tr}(\mathbf{E}_{\alpha\tilde{\mathbf{Z}}}(\gamma) - \mathbf{E}_{\tilde{\mathbf{X}}}(\gamma)) d\gamma \quad (31)$$

$$= \frac{1}{2} \int_0^{\text{snr}} q_{\mathbf{I}_n}(\tilde{\mathbf{X}}, \alpha^2, \gamma) d\gamma \quad (32)$$

where we have used the mmse function defined in (7) and the integral expression for the entropy function in [7].

Now, from (29) together with (26), it follows that

$$\lim_{\text{snr} \rightarrow \infty} \Delta I(\text{snr}) = 0 \quad (33)$$

which, from the integral expression in (32), further implies that the (smooth) integrand must have, at least, one zero crossing. However, from Theorem 1, we know that $q_{\mathbf{I}_n}(\tilde{\mathbf{X}}, \alpha^2, \gamma)$ can have, at most, one zero crossing. Consequently, in this case, $q_{\mathbf{I}_n}(\tilde{\mathbf{X}}, \alpha^2, \gamma)$ must have exactly one zero crossing. Also, from Theorem 1 and (33), we

can infer that there exists some $\text{snr}_0 \in (0, \infty)$ such that $q_{\mathbf{I}_n}(\tilde{\mathbf{X}}, \alpha^2, \gamma) < 0$, $\forall \gamma \in [0, \text{snr}_0)$, and $q_{\mathbf{I}_n}(\tilde{\mathbf{X}}, \alpha^2, \gamma) \geq 0$, $\forall \gamma \in [\text{snr}_0, \infty)$. Thus, it immediately follows that for finite snr , $\Delta I(\text{snr}) \leq 0$ and

$$\Delta I(\text{snr}) = I(\alpha \mathbf{Z}; \sqrt{\text{snr}} \alpha \mathbf{Z} + \mathbf{N}) - I(\mathbf{X}; \sqrt{\text{snr}} \mathbf{X} + \mathbf{N}) \quad (34)$$

$$= h(\sqrt{\text{snr}} \alpha \mathbf{Z} + \mathbf{N}) - h(\sqrt{\text{snr}} \mathbf{X} + \mathbf{N}) \leq 0. \quad (35)$$

It is now straightforward to see that

$$\exp\left(\frac{2}{n} h(\sqrt{\text{snr}} \mathbf{X} + \mathbf{N})\right) \geq \exp\left(\frac{2}{n} h(\sqrt{\text{snr}} \alpha \mathbf{Z} + \mathbf{N})\right) \quad (36)$$

$$= \exp\left(\frac{1}{n} \log((2\pi e)^n (\text{snr} \alpha^2 + 1)^n |\mathbf{R}_{\mathbf{N}}|)\right) \quad (37)$$

$$= (2\pi e)(\text{snr} \alpha^2 + 1) |\mathbf{R}_{\mathbf{N}}|^{\frac{1}{n}} \quad (38)$$

$$= \exp\left(\frac{2}{n} h(\sqrt{\text{snr}} \alpha \mathbf{Z})\right) + (2\pi e) |\mathbf{R}_{\mathbf{N}}|^{\frac{1}{n}} \quad (39)$$

$$= \exp\left(\frac{2}{n} h(\sqrt{\text{snr}} \mathbf{X})\right) + (2\pi e) |\mathbf{R}_{\mathbf{N}}|^{\frac{1}{n}} \quad (40)$$

which is exactly (25) up to scaling in $\sqrt{\text{snr}}$, which we can always take equal to 1. We note here that the I-MMSE relationship was used in [19] to prove Shannon's EPI, Costa's EPI and also the generalized EPI for linear transformations of a random vector.

III. FROM SCALAR TO VECTOR CHANNELS: DEFINITIONS AND PRELIMINARIES

In the previous section, we discussed the simple model presented in (13). We have shown that the “single crossing point” property initially proved for the scalar channel in [6] extends very smoothly and intuitively on to this model. The reason for the smooth transition is that, even though we are considering a multivariate scenario, all elements of the input vector undergo the same effect in the channel. They are all amplified by snr and distorted by additive standard Gaussian noise. From a more technical viewpoint, when one wants to search for a “single crossing point” property, one must define some scalar function of some scalar parameter, for which the property holds. In the model of (13) the intuitive choice is simply to take the trace of the MMSE as a function of snr . And indeed, this is just one possible linear combination included in Theorem 1, for which we have shown that the property can be extended.

Taking the next step, from this initial extension to the general model of (5), is a harder task. Moreover, there is no single method of doing so. In fact there are two degrees of freedom in this transition. First of all there is a need for some scalar parameter that will define \mathbf{H} . This parameter will be equivalent to the snr parameter in the scalar case or the simple model of (13). Secondly, there is a need for some scalar function of the matrix \mathbf{Q} . In the simple model of (13) we defined the function $q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma)$ which was simply taking some linear (positive semidefinite) combination of the elements of the matrix. The trace function is one example of such a combination, which is also the most intuitive extension; however, in the general model (or even the parallel model, which we will discuss shortly) the “single crossing point” property does not hold, in general, for the trace function. Thus, our goal is to find a “single crossing point” property that will be both elegant and, more importantly, useful and applicable.

As such, in this work we narrowed our investigation to the subset of parallel channels or diagonal matrices \mathbf{H} , for which we have the following result.

Lemma 4: For any two diagonal channel matrices, \mathbf{H}_1 and \mathbf{H}_2 , such that $\mathbf{0} \preceq \mathbf{H}_1 \preceq \mathbf{H}_2$, there exists a path $\mathbf{H}(t)$ such that the following holds:

- For all t , $\mathbf{H}(t) \succeq \mathbf{0}$ and is a diagonal matrix.
- For all t , $D_t \mathbf{H}(t) \succeq \mathbf{0}$ and is a diagonal matrix.
- $\mathbf{H}(0) = \mathbf{0}$.
- $\mathbf{H}(t_1) = \mathbf{H}_1$ and $\mathbf{H}(t_2) = \mathbf{H}_2$ where $0 \leq t_1 \leq t_2$.
- The diagonal elements of $\mathbf{H}(t)$ go to ∞ in a linear rate.

Proof: We need to define a function, $g_i(t)$, for each diagonal element of the matrix $\mathbf{H}(t)$. It suffices to choose any non-negative function $h_i(t)$ such that the area from 0 to t_1 will equal $[\mathbf{H}_1]_{ii}$ and the area from t_1 to t_2 will equal $[\mathbf{H}_2]_{ii} - [\mathbf{H}_1]_{ii}$. Given that, we can set the function to be $g_i(t) = \int_0^t h_i(\tau) d\tau$. The entire path, $\mathbf{H}(t)$, will be given by:

$$\mathbf{H}(t) = \text{Diag}(\{g_i(t)\}). \quad (41)$$

As required, this path passes between the zero matrix at $t = 0$, \mathbf{H}_1 at t_1 and \mathbf{H}_2 at t_2 . Since $h_i(t)$ are chosen nonnegative for all i we have a nonnegative and monotonically nondecreasing path for all t . The above construction guarantees that both $\mathbf{H}(t)$ and $D_t \mathbf{H}(t)$ will be diagonal matrices for all t . Moreover, we may also assume that the functions $h_i(t)$ plateau after complying with all other requirements, that is, from t_2 onwards. This assures that $g_i(t)$ goes to ∞ in a linear rate. ■

Note that the above lemma can be extended to M matrices $\mathbf{H}_j \preceq \mathbf{H}_{j+1}$ for $j = 1, \dots, M - 1$, using a similar construction.

Under the above detailed limitation, of restricting ourselves to parallel channels, we examine two different cases: phases two and three of our extension. In phase two, detailed in Section IV, we assume that the Gaussian covariance matrix defining the matrix \mathbf{Q} in (14) is that of a Gaussian distribution with independent elements, that is $\mathbf{R}_{\mathbf{X}_G} = \mathbf{\Lambda}_{\mathbf{X}_G}$ is a diagonal matrix. In this case we will see that the “single crossing point” property occurs for each and every diagonal element of \mathbf{Q} . This is not a straightforward extension of the scalar property, since the elements of the random input vector \mathbf{X} are, in general, not independent. In Section V we proceed to phase three where we allow any Gaussian distribution in the definition of \mathbf{Q} . In this phase we will see that the “single crossing point” property occurs for each and every eigenvalue of the matrix \mathbf{Q} . Surely, this is not a straightforward extension of any of the previous results. Moreover, the results of Section IV cannot be trivially deduced from the results of phase three, since restricting only the Gaussian covariance to be diagonal does not guarantee that the eigenvalues of \mathbf{Q} will be on its diagonal. Thus, we have two distinctive results. All results (including those of the previous section), fall back to the scalar “single crossing point” property result [6], [20] when both the arbitrary input vector \mathbf{X} and the Gaussian input random vector are restricted to have independent elements.

Before proceeding to examine these two cases we require a preliminary result. The basis for the applicability of

the “single crossing point” property in the scalar case and in the simple model of (5) is the I-MMSE relationship [7]. This is still the case in the extensions we are considering next, however, we require also an extension of the I-MMSE result which was derived by Palomar and Verdú in [17]:

$$\nabla_{\mathbf{H}} I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{N}) = \mathbf{H}\mathbf{E}. \quad (42)$$

This relationship was derived for complex-valued variables, however it holds verbatim for real-valued variables. Assuming the channel coefficients can be written as a function of a single parameter, t , we can rewrite the above relationship as an integral over this parameter, which results with the following expression:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}(t)) &= I(\mathbf{X}; \mathbf{H}(t)\mathbf{X} + \mathbf{N}) \\ &= \int_{\tau=0}^t \mathbf{1}^\top (\mathbf{H}(\tau)\mathbf{E}_{\mathbf{X}}(\tau) \circ \mathbf{D}_\tau \mathbf{H}(\tau)) \mathbf{1} \, d\tau \\ &= \int_{\tau=0}^t \text{Tr} \left((\mathbf{H}(\tau)\mathbf{E}_{\mathbf{X}}(\tau))^\top \mathbf{D}_\tau \mathbf{H}(\tau) \right) d\tau \end{aligned} \quad (43)$$

$$= \int_{\tau=0}^t \text{Tr} (\mathbf{B}(\tau)\mathbf{E}_{\mathbf{X}}(\tau)) \, d\tau \quad (44)$$

where we have used the following definition:

$$\mathbf{B}(t) \equiv \mathbf{H}(t) (\mathbf{D}_t \mathbf{H}(t))^\top. \quad (45)$$

This also carries over to the conditioned case as follows:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}(t)|U) &= I(\mathbf{X}; \mathbf{H}(t)\mathbf{X} + \mathbf{N}|U) \\ &= \int_{\tau=0}^t \text{Tr} \left((\mathbf{H}(\tau)\mathbf{E}_{\mathbf{X}|U}(\tau))^\top \mathbf{D}_\tau \mathbf{H}(\tau) \right) d\tau \end{aligned} \quad (46)$$

$$= \int_{\tau=0}^t \text{Tr} (\mathbf{B}(\tau)\mathbf{E}_{\mathbf{X}|U}(\tau)) \, d\tau. \quad (47)$$

IV. VECTOR CHANNEL: COMPARING WITH AN INDEPENDENT GAUSSIAN DISTRIBUTION

We begin our analysis of the extended model (5), limited to parallel channel matrices, by assuming that the Gaussian covariance matrix, defining the matrix \mathbf{Q} , is that of an independent distribution, that is, $\mathbf{R}_{\mathbf{X}_G} = \mathbf{\Lambda}_{\mathbf{X}_G}$, throughout this section. Recall, nonetheless, that \mathbf{X} remains completely arbitrary. More precisely, we consider the following matrix:

$$\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = \mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t) - \mathbf{E}_{\mathbf{X}}(t) \quad (48)$$

$$= \text{Diag} \left(\left\{ \frac{[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2 [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}} \right\} \right) - \mathbf{E}_{\mathbf{X}}(t). \quad (49)$$

Under these assumptions we will see, in Section IV-A, that a “single crossing point” property occurs for each and every diagonal element of the matrix \mathbf{Q} . After extending this result to the conditioned case, in Section IV-B, we will use the I-MMSE relationship, in Section IV-C, to show the effect of this property on information-theoretic quantities, and more specifically on the mutual information. Finally, in Section IV-D, we will put these results to use on a variant of the *degraded* BC, in order to show their applicability to information theory problems.

A. A Single Crossing Point Property on the Diagonal Elements of \mathbf{Q}

As pointed out above, our main result, in this section, is an extension of the “single crossing point” property. Precisely, we show that the property extends on each and every diagonal element of the matrix \mathbf{Q} . This result is given in the next theorem.

Theorem 2: The diagonal entries of the matrix-valued function $t \mapsto \mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$, defined in (48), have no nonnegative-to-negative zero crossings and, at most, a single negative-to-nonnegative zero crossing in the range $t \in [0, \infty)$. Moreover, let $t_0 \in [0, \infty)$ be the negative-to-nonnegative crossing point for $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$. Then,

- 1) $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, 0)]_{ii} \leq 0$.
- 2) $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ is a strictly increasing function in the range $t \in [0, t_0)$.
- 3) $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} \geq 0$ for all $t \in [\text{snr}_0, \infty)$.
- 4) Assuming $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$, we have that $\lim_{t \rightarrow \infty} [\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = 0$.
- 5) $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ is a continuous and monotonically increasing function in $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$.

Proof: Before giving the actual proof, let us first present an intermediate result.

Lemma 5: Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector such that $[\mathbf{R}_\mathbf{X}]_{ii} \leq [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$, where $i \in [1, n]$. Then, for every $t \geq 0$, we have

$$[\mathbf{E}_\mathbf{X}(t)]_{ii} \leq [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = \frac{[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2 [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}} \quad (50)$$

with equality if and only if $[\mathbf{X}]_i$ is Gaussian distributed, independent of the other entries of \mathbf{X} and such that $[\mathbf{R}_\mathbf{X}]_{ii} = [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$.

Proof: See Appendix A4. ■

Now, according to Lemma 5, for the case where $[\mathbf{R}_\mathbf{X}]_{ii} < [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$, the function $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ has no zeros and the statement in Theorem 1 is true. In addition, if $[\mathbf{X}]_i$ is Gaussian distributed (and independent of the other entries of the vector \mathbf{X}) with variance equal to $[\mathbf{R}_\mathbf{X}]_{ii} = [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$, then $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = 0$, $\forall t$, which also fulfills Theorem 1.

Thus, from this point, we can assume that $[\mathbf{R}_\mathbf{X}]_{ii} \geq [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$ and that $[\mathbf{X}]_i$ is not Gaussian distributed, independent of the other entries of \mathbf{X} , and with $[\mathbf{R}_\mathbf{X}]_{ii} = [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$. Now, for $t = 0$ we have $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, 0)]_{ii} = [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} - [\mathbf{R}_\mathbf{X}]_{ii} \leq 0$ as required.

Similarly as it was done in the proof of Theorem 1, in order to prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ can occur, we only need to show that the derivative of $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ with respect to t is positive for all values of t for which $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < 0$. Observe that $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < 0$ implies

$$\frac{[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2 [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}} = [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < [\mathbf{E}_\mathbf{X}(t)]_{ii}. \quad (51)$$

Now, from (48), it is clear that, in order to compute the derivative of $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$, we first need the derivative of $[\mathbf{E}_{\mathbf{X}}(t)]_{ii}$:

$$D_t[\mathbf{E}_{\mathbf{X}}(t)]_{ii} = D_{\mathbf{H}(t)}[\mathbf{E}_{\mathbf{X}}(t)]_{ii} D_t \mathbf{H}(t) \quad (52)$$

$$= \sum_{j=1}^n D_{[\mathbf{H}(t)]_{jj}} [\mathbf{E}_{\mathbf{X}}(t)]_{ii} D_t [\mathbf{H}(t)]_{jj} \quad (53)$$

where, in the last step, we have used the assumption that $\mathbf{H}(t)$ is a diagonal matrix for all t . From [14, Eq. (131)], we have

$$D_{[\mathbf{H}(t)]_{jj}} [\mathbf{E}_{\mathbf{X}}(t)]_{ii} = -2\mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ij} [\Phi_{\mathbf{X}}(\mathbf{Y}) \mathbf{H}(t)^T]_{ij}\} \quad (54)$$

$$= -2[\mathbf{H}(t)]_{jj} \mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ij}^2\}. \quad (55)$$

Recalling the definition $[\mathbf{B}(t)]_{ii} = [\mathbf{H}(t)]_{ii} D_t [\mathbf{H}(t)]_{ii}$ in (45), we are now ready to compute the derivative of $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$, which reads as

$$D_t[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} \quad (56)$$

$$= 2 \sum_{j=1}^n [\mathbf{B}(t)]_{jj} (\mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ij}^2\} - [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ij}^2) \quad (57)$$

$$= 2[\mathbf{B}(t)]_{ii} (\mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ii}^2\} - [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}^2) + 2 \sum_{j \neq i} [\mathbf{B}(t)]_{jj} \mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ij}^2\} \quad (58)$$

$$\geq 2[\mathbf{B}(t)]_{ii} (\mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ii}^2\} - [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}^2) \quad (59)$$

$$> 2[\mathbf{B}(t)]_{ii} (\mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ii}^2\} - (\mathbb{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y})]_{ii}\})^2) \quad (60)$$

$$\geq 0 \quad (61)$$

where (57) follows from the fact that for Gaussian input distributions (not necessarily i.i.d.), the conditional MMSE matrix $\Phi_{\mathbf{X}_G}(\mathbf{y})$ does not depend on the observation \mathbf{y} , i.e., $\mathbf{E}_G(\mathbf{R}_{\mathbf{X}_G}, t) = \Phi_{\mathbf{X}_G}$. Equation (58) is due to the fact that the entries of the Gaussian input distribution \mathbf{X}_G are independent and, thus, its MMSE matrix is diagonal; (59) is due to the fact that $[\mathbf{B}(t)]_{ii} \geq 0$, as shown in Lemma 4; (60) follows from the assumption $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < 0$ and (61) can be derived from Jensen's inequality.

Observe that the inequality in (61), which holds for values of t such that $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < 0$, also proves the second item in Theorem 2 and the third one follows directly from the inexistence of nonnegative-to-negative zero crossings. Regarding the fourth item, it is clear that $\lim_{t \rightarrow \infty} [\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = 0$, as both terms in the expression of $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ in (49) tend to zero, when $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$. Finally, the last property is a direct consequence of the definition of the function $\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ (49). ■

We now define the following function:

$$d_i(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = [\mathbf{B}(t)]_{ii} [\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} \quad (62)$$

and also,

$$d(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = \sum_{i=1}^n d_i(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \quad (63)$$

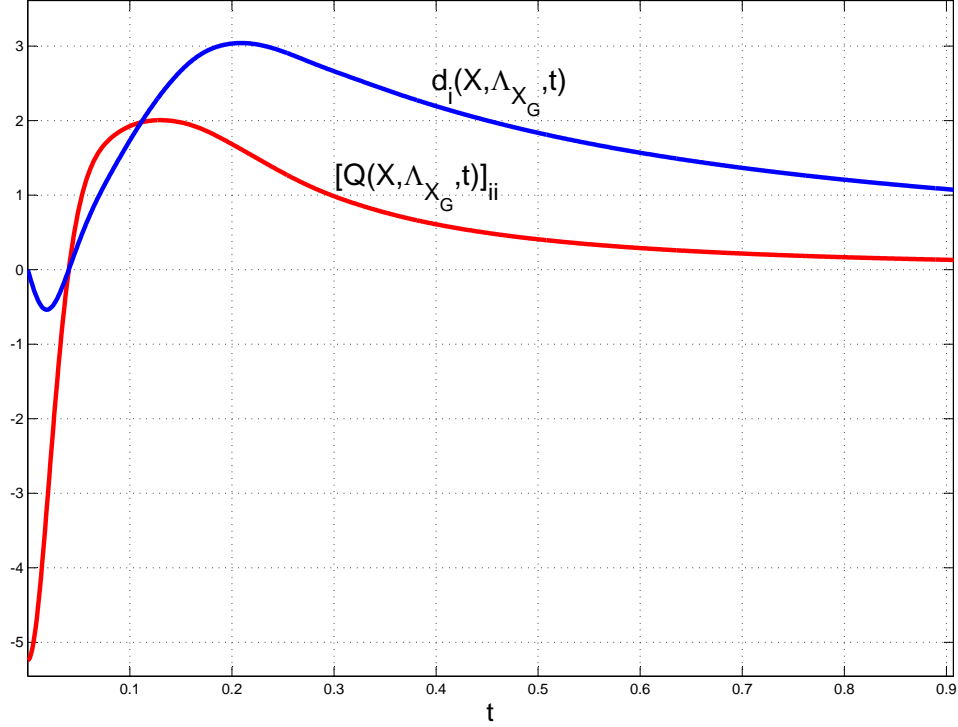


Fig. 1. An example of the function $[Q(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t)]_{ii}$ (in red) and the matching function $d_i(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t)$ (in blue). Both have the same single negative-to-nonnegative zero crossing in the range $t \in (0, \infty)$.

For which we can give the following two corollaries,

Corollary 1: Let $\mathbf{X} \in \mathbb{R}^n$ be any random vector. The function $d_i(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t)$ has the following properties:

- 1) $d_i(\mathbf{X}, \Lambda_{\mathbf{X}_G}, 0) = 0$.
- 2) It has at most a single negative-zero-positive crossing in the range $t \in (0, \infty)$.
- 3) When $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$ we have that, $\lim_{t \rightarrow \infty} d_i(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t) = 0$.
- 4) If $[\Lambda_{\mathbf{X}_G}]_{ii} = [\mathbf{R}_\mathbf{X}]_{ii}$, then $d_i(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t) \geq 0$ for all t . Furthermore, $d_i(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t)$ is a continuous and monotonically increasing function in $[\Lambda_{\mathbf{X}_G}]_{ii}$.

Proof: The first three properties follow from Theorem 2 and the fact that $[\mathbf{B}(t)]_{ii}$ is zero at $t = 0$, non-negative for all other values of $t \in (0, \infty)$ and $[\mathbf{B}(t)]_{ii}$ goes to ∞ in a linear rate, as shown in Lemma 4. The fourth property is a direct result of Lemma 5 and the fifth item of Theorem 2. ■

Figure 1 illustrates this property, in which the negative-zero-positive crossing of $d_i(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t)$ is simply a negative-to-nonnegative zero crossing and, thus, agrees with the negative-to-nonnegative zero crossing of $[Q(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t)]_{ii}$.

Corollary 2: Let $\mathbf{X} \in \mathbb{R}^n$ be any random vector. The function $d(\mathbf{X}, \Lambda_{\mathbf{X}_G}, t)$ is either negative for all t , or

there exists $t' \in [0, \infty)$ such that for all $t > t'$ the function $d(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ is nonnegative. Moreover, when $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$ we have that, $\lim_{t \rightarrow \infty} d(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = 0$, and if $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} = [\mathbf{R}_{\mathbf{X}}]_{ii}$ for all i , then $d(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \geq 0$ for all t .

B. The Conditioned Case

Before proceeding to understanding the implications of the above results on information-theoretic quantities, we would like to extend these results to the conditioned case.

Let us begin with the conditioned MMSE matrix. We first consider the following matrix quantity:

$$\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t, \mathbf{u}) = \mathbb{E}\{(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{H}(t)\mathbf{X} + \mathbf{N}, \mathbf{U} = \mathbf{u}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{H}(t)\mathbf{X} + \mathbf{N}, \mathbf{U} = \mathbf{u}\})^\top | \mathbf{U} = \mathbf{u}\} \quad (64)$$

$$= \mathbb{E}\{(\mathbf{X}_{\mathbf{u}} - \mathbb{E}\{\mathbf{X}_{\mathbf{u}} | \mathbf{H}(t)\mathbf{X}_{\mathbf{u}} + \mathbf{N}\})(\mathbf{X}_{\mathbf{u}} - \mathbb{E}\{\mathbf{X}_{\mathbf{u}} | \mathbf{H}(t)\mathbf{X}_{\mathbf{u}} + \mathbf{N}\})^\top\} \quad (65)$$

where $\mathbf{X}_{\mathbf{u}}$ is a random vector distributed according to $P_{\mathbf{X}|\mathbf{U}=\mathbf{u}}$. The conditioned MMSE matrix is simply the expectation of (64) according to the distribution of the random vector \mathbf{U} :

$$\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t) = \mathbb{E}\{\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t, \mathbf{U})\} = \mathbb{E}\{(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{H}(t)\mathbf{X} + \mathbf{N}, \mathbf{U}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{H}(t)\mathbf{X} + \mathbf{N}, \mathbf{U}\})^\top\}. \quad (66)$$

Another important quantity that needs to be extended to the conditioned case is:

$$\Phi_{\mathbf{X}_{\mathbf{u}}}(\mathbf{y}) = \mathbb{E}\{(\mathbf{X}_{\mathbf{u}} - \mathbb{E}\{\mathbf{X}_{\mathbf{u}} | \mathbf{y}\})(\mathbf{X}_{\mathbf{u}} - \mathbb{E}\{\mathbf{X}_{\mathbf{u}} | \mathbf{y}\})^\top | \mathbf{y}\} \quad (67)$$

$$= \mathbb{E}\{(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{y}, \mathbf{U} = \mathbf{u}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X} | \mathbf{y}, \mathbf{U} = \mathbf{u}\})^\top | \mathbf{y}, \mathbf{U} = \mathbf{u}\} \quad (68)$$

$$= \Phi_{\mathbf{X}}(\mathbf{y}, \mathbf{U} = \mathbf{u}) \quad (69)$$

where, as in the unconditioned case, this function, in general, depends on both \mathbf{u} and \mathbf{y} , thus, we have $\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t, \mathbf{u}) = \mathbb{E}\{\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U} = \mathbf{u})\}$, where the expectation is over \mathbf{Y} . However, when the input distribution of $\mathbf{X}_{\mathbf{u}}$ is Gaussian $\Phi_{\mathbf{X}}(\mathbf{y}, \mathbf{U} = \mathbf{u})$ is independent of \mathbf{y} . In a similar manner, we have the following:

$$\mathbf{Q}(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = \mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t) - \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t, \mathbf{u}) \quad (70)$$

and, thus, we also have:

$$\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = \mathbb{E}_{\mathbf{U}}\{\mathbf{Q}(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)\} = \mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t) - \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t). \quad (71)$$

Using these definitions we can now extend the results of Theorem 2 to the conditioned case in the following theorem.

Theorem 3: Let $\mathbf{U} - \mathbf{X} - \mathbf{Y}$ form a Markov chain. Then, the diagonal entries of the matrix-valued function $t \mapsto \mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$, defined in (71), have no nonnegative-to-negative zero crossings and, at most, a single negative-to-nonnegative zero crossing in the range $t \in [0, \infty)$. Moreover, let $t_0 \in [0, \infty)$ be the negative-to-nonnegative crossing point for $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$. Then,

- 1) $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, 0)]_{ii} \leq 0$.
- 2) $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ is a strictly increasing function in the range $t \in [0, t_0)$.

- 3) $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} \geq 0$ for all $t \in [\text{snr}_0, \infty)$.
- 4) When $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$ we have that $\lim_{t \rightarrow \infty} [\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = 0$.
- 5) $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ is a continuous and monotonically increasing function in $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$.

Proof: If $[\mathbf{X}]_i$ is Gaussian distributed (independent of \mathbf{U} and independent of the other entries of the vector \mathbf{X}) with variance equal to $[\mathbf{R}_\mathbf{X}]_{ii} = [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$, then $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = 0, \forall t$, which also fulfills Theorem 3. Thus, from this point, we can assume that $[\mathbf{X}]_i$ is not “Gaussian distributed, independent of \mathbf{U} and independent of the other entries of \mathbf{X} , and such that $[\mathbf{R}_\mathbf{X}]_{ii} = [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$ ”.

In this conditioned case, it is harder to determine, up front, all cases in which the function $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ has no zeros. Thus, contrary to the approach used in the proof of Theorem 2, we first prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ can occur. The first property is a direct consequence of this, and there is no need to determine the exact conditions under which the function has no zeros. This approach could have also been used in proving Theorem 2, however in the unconditioned case we can easily determine the set of cases in which $[\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ has no zeros.

Similarly to the proof of Theorem 2, in order to prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ can occur, we only need to show that the derivative of $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ with respect to t is positive for all values of t for which $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < 0$. According to equations (57) and (59) we have the following lower bound:

$$D_t[\mathbf{Q}(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = 2 \sum_{j=1}^n [\mathbf{B}(t)]_{jj} (E\{[\Phi_{\mathbf{X}_u}(\mathbf{Y})]_{ij}^2\} - [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ij}^2) \quad (72)$$

$$\geq 2[\mathbf{B}(t)]_{ii} (E\{[\Phi_{\mathbf{X}_u}(\mathbf{Y})]_{ii}^2\} - [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}^2). \quad (73)$$

Now we can take expectation over \mathbf{U} on both sides and attain the following:

$$D_t[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = E_{\mathbf{U}}\{D_t[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}\} \quad (74)$$

$$\geq E_{\mathbf{U}}\{2[\mathbf{B}(t)]_{ii} (E\{[\Phi_{\mathbf{X}_u}(\mathbf{Y})]_{ii}^2\} - [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}^2)\} \quad (75)$$

$$= 2[\mathbf{B}(t)]_{ii} (E\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_{ii}^2\} - [\mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}^2) \quad (76)$$

$$> 2[\mathbf{B}(t)]_{ii} (E\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_{ii}^2\} - (E\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_{ii}\})^2) \quad (77)$$

$$\geq 0 \quad (78)$$

where (75) is due to (73), (77) follows from the assumption $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < 0$ and (78) can be derived from Jensen's inequality.

Observe that the inequality in (77), which holds for values of t such that $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} < 0$, also proves the second item in Theorem 3 and the third one follows directly from the inexistence of nonnegative-to-negative zero crossings. Regarding the fourth item, it is clear that $\lim_{t \rightarrow \infty} [\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} = 0$, as both terms in the expression of $[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii}$ in (71) tend to zero, when $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$. Finally, the last property is a direct consequence of the definition of the function $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ in (71). ■

We now extend the definition of the function $d_i(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ (62) and the function $d(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ (63) to the conditioned case:

$$d_i(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = [\mathbf{B}(t)]_{ii}[\mathbf{Q}(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} \quad (79)$$

$$d(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = [\mathbf{B}(t)]_{ii}[\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)]_{ii} \quad (80)$$

and also,

$$d(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = \sum_{i=1}^n d_i(\mathbf{X}|\mathbf{U} = \mathbf{u}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \quad (81)$$

$$d(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = \sum_{i=1}^n d_i(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \quad (82)$$

For which we can extend corollaries 1 and 2 as follows,

Corollary 3: Let $\mathbf{U} - \mathbf{X} - \mathbf{Y}$ form a Markov chain such that the random vector $\mathbf{X}|\mathbf{U} = \mathbf{u} \in \mathbb{R}^n$ has covariance matrix $\mathbf{R}_{\mathbf{X}|\mathbf{U}=\mathbf{u}}$. The function $d_i(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ has the following properties:

- 1) $d_i(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, 0) = 0$
- 2) It has at most a single negative-zero-positive crossing in the range $t \in (0, \infty)$.
- 3) When $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$ we have that, $\lim_{t \rightarrow \infty} d_i(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = 0$.
- 4) If $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} = [\mathbf{R}_{\mathbf{X}}]_{ii}$, then $d_i(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \geq 0$ for all t . Furthermore, $d_i(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ is a continuous and monotonically increasing function in $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$.

Proof: The first three properties follow directly from Theorem 3 and the fact that $[\mathbf{B}(t)]_{ii}$ is zero at $t = 0$ and nonnegative for all other values of $t \in (0, \infty)$ and $[\mathbf{B}(t)]_{ii}$ goes to ∞ in a linear rate, as shown in Lemma 4. The fourth property is a direct result of Lemma 5, and the fifth property in Theorem 3. ■

Corollary 4: Let $\mathbf{U} - \mathbf{X} - \mathbf{Y}$ form a Markov chain. The function $d(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ is either negative for all t , or there exists $t' \in [0, \infty)$ such that for all $t > t'$ the function $d(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ is nonnegative. Moreover, when $\lim_{t \rightarrow \infty} [\mathbf{H}(t)]_{ii} = \infty$ we have that, $\lim_{t \rightarrow \infty} d_i(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = 0$, and if $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} = [\mathbf{R}_{\mathbf{X}}]_{ii}$ for all i , then $d(\mathbf{X}|\mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \geq 0$ for all t .

C. Properties of the Mutual Information

So far, we have seen properties of the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ or, more precisely, of its diagonal elements. We have seen that these properties extend naturally to the conditioned case, and also to the function $d_i(\mathbf{X}, \mathbf{\Lambda}_{\mathbf{X}_G}, t)$ and its conditioned version. In this section, our goal is to use these results to derive new properties on the mutual information between the input and the output of parallel Gaussian channels. In order to derive these results we put to use the I-MMSE relationship, as given in equations (43)-(44) and (46)-(47).

For the sake of compactness we will write the properties in this section only for the more general, conditioned case, from which one can easily derive the respective unconditioned theorems.

Theorem 4: Let $U - X - Y$ form a Markov chain. Assume an independent Gaussian input, X_G , with covariance Λ_{X_G} , such that for all i ,

$$I([X]_i; [Y(t_e)]_i | U) = I([X_G]_i; [Y_G(t_e)]_i) \quad (83)$$

where

$$Y(t_e) = H(t_e)X + N \quad \text{and} \quad (84)$$

$$Y_G(t_e) = H(t_e)X_G + N. \quad (85)$$

Then $d(X|U, \Lambda_{X_G}, t) \geq 0$ for all $t \geq t_e$.

Proof: Let us define $X^{\text{ind cond.}} \in \mathbb{R}^n$ as a random vector with independent elements when conditioned on U , and with distribution of each pair $([X^{\text{ind cond.}}]_i, U)$ being the same as the marginal distribution of the corresponding pair $([X]_i, U)$. Thus, $[E_{X^{\text{ind cond.}}|U}]_{ii}$ is basically the MMSE of $[X^{\text{ind cond.}}]_i$ from U and $[Y(t)]_i$, which is:

$$[Y(t)]_i = [H(t)X^{\text{ind cond.}} + N]_i = [H(t)]_{ii} [X^{\text{ind cond.}}]_i + [N]_i \quad (86)$$

where the equality holds due to the fact that the channel matrix $H(t)$ is diagonal for all t and N is standard Gaussian. Using these definitions we can give the following special case of (47):

$$I([X]_i; [Y(t)]_i | U) = I([X]_i; [H(t)]_{ii} [X^{\text{ind cond.}}]_i + [N]_i | U) \quad (87)$$

$$= \int_{\tau=0}^t [H(\tau)]_{ii} [E_{X^{\text{ind cond.}}|U}(\tau)]_{ii} [D_\tau H(\tau)]_{ii} d\tau \quad (88)$$

$$= \int_{\tau=0}^t [B(\tau)]_{ii} [E_{X^{\text{ind cond.}}|U}(\tau)]_{ii} d\tau. \quad (89)$$

Putting this together with the assumption, we have,

$$0 = I([X_G]_i; [Y_G(t_e)]_i) - I([X]_i; [Y(t_e)]_i | U) = \int_{\tau=0}^{t_e} d_i(X^{\text{ind cond.}}|U, \Lambda_{X_G}, \tau) d\tau. \quad (90)$$

Now, due to Corollary 4 we can conclude that there exists a $t_0 \in [0, t_e]$ such that $d_i(X^{\text{ind cond.}}|U, \Lambda_{X_G}, t) \geq 0$ for all $t > t_0$ and as a result, $d_i(X^{\text{ind cond.}}|U, \Lambda_{X_G}, t) \geq 0$ for all $t > t_e$. Now, for all t we have that $[E_{X|U}(t)]_{ii} \leq [E_{X^{\text{ind cond.}}|U}(t)]_{ii}$. Thus, if the negative-zero-positive crossing of $d_i(X^{\text{ind cond.}}|U, \Lambda_{X_G}, t)$ is at t_0 , the negative-zero-positive crossing of $d_i(X|U, \Lambda_{X_G}, t)$ is at a $t'_0 \leq t_0$. From this we can conclude that also $d_i(X|U, \Lambda_{X_G}, t) \geq 0$ for all $t > t_e$. Finally, since this holds for every i , it also holds for the summation over i , i.e., for the function $d(X|U, \Lambda_{X_G}, t)$, concluding the proof. ■

We are now ready to give the main theorem of this section.

Theorem 5: Let $U - X - Y$ form a Markov chain. For any $t_e \in [0, \infty)$, there exists an independent Gaussian input, X_G , with covariance Λ_{X_G} such that the following properties hold:

- 1) $d(X|U, \Lambda_{X_G}, t) \geq 0$ for all $t \geq t_e$.
- 2) $I(X; Y(t_e)|U) = I(X_G; Y_G(t_e))$, where $Y(t_e)$ and $Y_G(t_e)$ are as defined in (84) and (85) respectively.
- 3) $[\Lambda_{X_G}]_{ii} \leq [R_X]_{ii}$ for all i .

Proof: We provide a constructive proof, and show how one can build an independent Gaussian input distribution complying with all three requirements. We begin by examining the meaning of the second requirement. First, recall the I-MMSE relationship in the parallel setting, given in equation (47),

$$I(\mathbf{X}; \mathbf{Y}(t_e)|U) = \int_{\tau=0}^{t_e} \text{Tr}(\mathbf{B}(\tau)\mathbf{E}_{\mathbf{X}|U}(\tau)) d\tau. \quad (91)$$

Now, the second requirement is equivalent to the following equality,

$$\begin{aligned} 0 &= I(\mathbf{X}_G; \mathbf{Y}_G(t_e)) - I(\mathbf{X}; \mathbf{Y}(t_e)|U) = \int_{\tau=0}^{t_e} \text{Tr}(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau)) d\tau \\ &= \int_{\tau=0}^{t_e} \sum_{i=1}^n d_i(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau \\ &= \sum_{i=1}^n \int_{\tau=0}^{t_e} d_i(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau. \end{aligned} \quad (92)$$

Thus, we wish to show the existence of an independent Gaussian input distribution which complies with requirements 1, 3 and (92). There are different ways to attain equality in (92), however since we need only to show the existence of a specific independent Gaussian distribution, we follow one possible approach, which is to require the following,

$$\int_{\tau=0}^{t_e} d_i(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau = 0, \quad \forall i. \quad (93)$$

Now, according to the fourth property in Corollary 3 we know that,

$$d_i(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \geq 0 \quad (94)$$

for all t , when $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} = [\mathbf{R}_{\mathbf{X}}]_{ii}$, and that it is continuous and monotonically increasing in the value of $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii}$ (and trivially negative, for all t , when $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} = 0$). Thus, there exists a number $\eta_i \in [0, 1]$ such that setting $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} = \eta_i[\mathbf{R}_{\mathbf{X}}]_{ii}$ results with the equality in (93). Due to the second property in Corollary 3, we know that either $d_i(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, t) = 0$ for all t or that there exists a single negative-zero-positive crossing in the range $[0, t_e]$. In both cases the setting $[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} = \eta_i[\mathbf{R}_{\mathbf{X}}]_{ii}$ results with $d_i(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \geq 0$ for all $t > t_e$. Since there exists such an η_i for every i we comply also with requirements 1 and 3, and conclude the proof. ■

Remark 4: Note that the above choice of $\mathbf{\Lambda}_{\mathbf{X}_G}$ does not necessarily imply $\mathbf{\Lambda}_{\mathbf{X}_G} \preceq \mathbf{R}_{\mathbf{X}}$. However, we can conclude that $\mathbf{\Lambda}_{\mathbf{X}_G} \not\preceq \mathbf{R}_{\mathbf{X}}$.

The following is a simple corollary of the above theorem.

Corollary 5: Given any arbitrary *independent* input distribution over $\mathbf{X} \in \mathbb{R}^n$, with covariance $\mathbf{\Lambda}_{\mathbf{X}}$, and any t_e , there exists an independent Gaussian input, \mathbf{X}_G , with covariance $\mathbf{\Lambda}_{\mathbf{X}_G}$ such that

$$I(\mathbf{X}; \mathbf{H}(t_e)\mathbf{X} + \mathbf{N}) = I(\mathbf{X}_G; \mathbf{H}(t_e)\mathbf{X}_G + \mathbf{N}) \quad (95)$$

$$\mathbf{\Lambda}_{\mathbf{X}_G} \preceq \mathbf{\Lambda}_{\mathbf{X}} \quad (96)$$

$$\text{and } \mathbf{E}_G(\mathbf{\Lambda}_{\mathbf{X}_G}, t_e) \preceq \mathbf{E}_{\mathbf{X}}(t_e) \quad (97)$$

D. Application: The Degraded Parallel Gaussian BC Capacity Region under Per-antenna Power Constraint

We now show that Theorem 5 can be used in providing a converse proof for the *degraded* parallel Gaussian BC capacity region under a per-antenna power constraint. We consider the following model,

$$\begin{aligned} \mathbf{Y}_1[m] &= \mathbf{H}_1 \mathbf{X}[m] + \mathbf{N}_1[m] \\ \mathbf{Y}_2[m] &= \mathbf{H}_2 \mathbf{X}[m] + \mathbf{N}_2[m] \end{aligned} \quad (98)$$

where $\mathbf{N}_1[m]$ and $\mathbf{N}_2[m]$ are standard additive Gaussian noise vectors independent for different time indices m , and \mathbf{H}_1 and \mathbf{H}_2 are diagonal positive semidefinite matrices such that $\mathbf{H}_1 \preceq \mathbf{H}_2$. $\mathbf{X} \in \mathbb{R}^n$ is the random input vector, and it is assumed independent for different time indices m . Note that m is the time index and should not be confused with the scalar parameter t which is used as a “MIMO snr parameter”, i.e., the parameter t determines the channel matrix $\mathbf{H}(t)$.

We consider a per-antenna power constraint:

$$[\mathbb{E}\{\mathbf{X}\mathbf{X}^\top\}]_{ii} \leq P_i \quad \forall i, 1 \leq i \leq n. \quad (99)$$

Since we have a *degraded* BC, we can use the single-letter expression given in [21],

$$\begin{aligned} R_1 &\leq I(\mathbf{U}; \mathbf{Y}_1) \\ R_2 &\leq I(\mathbf{X}; \mathbf{Y}_2 | \mathbf{U}) \end{aligned} \quad (100)$$

where \mathbf{U} is an auxiliary random vector over a certain alphabet that satisfies the Markov relation $\mathbf{U} - \mathbf{X} - (\mathbf{Y}_1, \mathbf{Y}_2)$. The following proof was originally given for the scalar Gaussian BC in [6], [20] and we now extend it to the *degraded* parallel Gaussian channel. Using Lemma 4 we can construct a path such that:

$$\begin{aligned} \mathbf{H}(t_2) &= \mathbf{H}_2 \\ \mathbf{H}(t_1) &= \mathbf{H}_1 \\ \mathbf{H}(0) &= \mathbf{0} \end{aligned} \quad (101)$$

where $0 \leq t_1 \leq t_2$ and $\mathbf{H}(t)$ is diagonal for all $t \in [0, t_2]$.

Now, assume a pair (\mathbf{U}, \mathbf{X}) such that \mathbf{X} has covariance $\mathbf{R}_\mathbf{X}$. According to Theorem 5, there exists an independent Gaussian vector, \mathbf{X}_G , with covariance matrix $\mathbf{\Lambda}_{\mathbf{X}_G}$ such that the following properties hold:

$$I(\mathbf{X}; \mathbf{Y}_1 | \mathbf{U}) = I(\mathbf{X}; \mathbf{H}(t_1)\mathbf{X} + \mathbf{N} | \mathbf{U}) = I(\mathbf{X}_G; \mathbf{Y}_G(t_1)) = I(\mathbf{X}_G; \mathbf{H}(t_1)\mathbf{X}_G + \mathbf{N}) \quad (102)$$

$$d(\mathbf{X} | \mathbf{U}, \mathbf{\Lambda}_{\mathbf{X}_G}, t) \geq 0, \quad \forall t \geq t_1 \quad (103)$$

$$[\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} \leq [\mathbf{R}_\mathbf{X}]_{ii} \quad \forall i. \quad (104)$$

Using the I-MMSE relationship (47) we can write,

$$I(\mathbf{X}_G; \mathbf{H}(t)\mathbf{X}_G + \mathbf{N}) - I(\mathbf{X}; \mathbf{H}(t)\mathbf{X} + \mathbf{N}|U) = \int_{\tau=0}^t \text{Tr}(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau)) d\tau \quad (105)$$

$$= \int_{\tau=0}^t \sum_{i=1}^n d_i(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau \quad (106)$$

$$= \int_{\tau=0}^t d(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau. \quad (107)$$

Using the above properties on (107) we have that for any $t' > t_1$,

$$I(\mathbf{X}_G; \mathbf{H}(t')\mathbf{X}_G + \mathbf{N}) - I(\mathbf{X}; \mathbf{H}(t')\mathbf{X} + \mathbf{N}|U) = \int_{\tau=0}^{t_1} d(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau + \int_{\tau=t_1}^{t'} d(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau \quad (108)$$

$$= 0 + \int_{\tau=t_1}^{t'} d(\mathbf{X}|U, \mathbf{\Lambda}_{\mathbf{X}_G}, \tau) d\tau \geq 0 \quad (109)$$

where the second transition is due to (102) and the inequality is due to (103). Thus, we have shown the existence of an independent Gaussian vector, \mathbf{X}_G , with covariance matrix $\mathbf{\Lambda}_{\mathbf{X}_G}$, with the following properties:

$$I(\mathbf{X}; \mathbf{H}_1\mathbf{X} + \mathbf{N}|U) = \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1\mathbf{\Lambda}_{\mathbf{X}_G}\mathbf{H}_1^\top| \quad (110)$$

$$I(\mathbf{X}; \mathbf{H}_2\mathbf{X} + \mathbf{N}|U) \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_2\mathbf{\Lambda}_{\mathbf{X}_G}\mathbf{H}_2^\top| \quad (111)$$

$$\text{and } [\mathbf{\Lambda}_{\mathbf{X}_G}]_{ii} \leq [\mathbf{R}_\mathbf{X}]_{ii} \quad \forall i. \quad (112)$$

Using these properties on the single-letter expression (100) we obtain the following outer bound:

$$\begin{aligned} R_1 &\leq I(\mathbf{U}; \mathbf{Y}_1) = I(\mathbf{X}; \mathbf{Y}_1) - I(\mathbf{X}; \mathbf{Y}_1|U) \\ &\leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1\mathbf{P}\mathbf{H}_1^\top| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1\mathbf{\Lambda}_{\mathbf{X}_G}\mathbf{H}_1^\top| = \frac{1}{2} \log \frac{|\mathbf{I} + \mathbf{H}_1\mathbf{P}\mathbf{H}_1^\top|}{|\mathbf{I} + \mathbf{H}_1\mathbf{\Lambda}_{\mathbf{X}_G}\mathbf{H}_1^\top|} \end{aligned} \quad (113)$$

$$R_2 \leq I(\mathbf{X}; \mathbf{Y}_2|U) \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_2\mathbf{\Lambda}_{\mathbf{X}_G}\mathbf{H}_2^\top| \quad (114)$$

where \mathbf{P} is a diagonal matrix with $[\mathbf{P}]_{ii} = P_i$ for all i . This outer bound is tight and the achievability is well-known using superposition coding. This approach can be extended to the M-user scenario as shown in Appendix B.

V. VECTOR CHANNEL: COMPARING WITH A GENERAL GAUSSIAN DISTRIBUTION

In this section we extend our analysis of the previous section. We continue looking into the model given in (5), limited to parallel channel matrices, however we now allow the Gaussian covariance matrix, defining the matrix \mathbf{Q} , to be any proper covariance matrix. In other words, we no longer limit ourselves to independent Gaussian inputs. For this, more general setting, we will see in Section V-A that a “single crossing point” property occurs for each and every eigenvalue of the matrix \mathbf{Q} . After extending this result to the conditioned case, in Section V-B, we will use the I-MMSE relationship, in Section V-C, to show the effect of this property on information-theoretic quantities, and more specifically on the mutual information. We will relate these results to the Fisher information in Section V-D. Finally, in Sections V-E and V-F we will put these results to use in the *degraded* BC capacity converse proof, for both the compound and non-compound scenarios.

A. Single Crossing Point for Each Eigenvalue of $\mathbf{Q}(t)$

In this section we prove the main result of this paper: showing that each eigenvalue of the matrix \mathbf{Q} has at most a single negative-to-nonnegative zero crossing. This is, to our understanding, not an intuitive extension of the “single crossing point” property, which emphasizes the importance of the eigenvalues in the analysis of MIMO scenarios.

For the proof of the main theorem, we require the following lemma, which might also be of interest on its own.

Lemma 6: The following lower bound holds:

$$D_t \mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t) \succeq 2 (\mathbf{E}_X(t) \mathbf{B}(t) \mathbf{E}_X^\top(t) - \mathbf{E}_G(t) \mathbf{B}(t) \mathbf{E}_G^\top(t)) \quad (115)$$

where $\mathbf{B}(t)$ was defined in (45) and assumed a positive semidefinite diagonal matrix for all t (see Lemma 4).

Proof: See Appendix A5. ■

We are now ready to proceed to the main result of the paper:

Theorem 6: Each eigenvalue of $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ has, *at most*, a single negative-to-nonnegative zero crossing of the horizontal axis.

Proof: Loosely speaking, the proof is based on proving that, once an eigenvalue has become (or is) nonnegative, it cannot become negative. Thus, from the (weak) continuity of the eigenvalues as a function of t , that follows from [22, App. D], the eigenvalues can cross the horizontal axis, at most, once. Also from continuity arguments, it is easy to see that we must limit our study of the eigenvalues of $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ to the values of t where the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ is singular (i.e., a subset of its eigenvalues are zero) as it is the only possible situation where a zero crossing can occur. Finally, throughout this proof and for the sake of simplicity we will use the simplified notation $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t) = \mathbf{Q}(t) \triangleq \mathbf{E}_G(t) - \mathbf{E}_X(t)$ because the entire proof is given for any constant setting of the input random vector \mathbf{X} and the Gaussian covariance $\mathbf{R}_{\mathbf{X}_G}$.

We begin by stating a few supporting results and giving some preliminary definitions.

Lemma 7: Let \mathbf{A} and \mathbf{B} be two n -dimensional positive semidefinite matrices, i.e., $\mathbf{A} \succeq \mathbf{0}$, $\mathbf{B} \succeq \mathbf{0}$. Then, there exists an invertible matrix \mathbf{S} such that both \mathbf{SAS}^T and \mathbf{SBS}^T are diagonal matrices.

Proof: See Appendix A6. ■

Let us consider the simultaneous decomposition of $\{\mathbf{E}_G(t), \mathbf{E}_X(t)\}$ according to Lemma 7 as:

$$\mathbf{E}_G(t) = \mathbf{V}(t)^\top \boldsymbol{\Sigma}_G(t) \mathbf{V}(t) \quad (116)$$

$$\mathbf{E}_X(t) = \mathbf{V}(t)^\top \boldsymbol{\Sigma}_X(t) \mathbf{V}(t)$$

where $\mathbf{V}(t)$ is an invertible matrix and $\boldsymbol{\Sigma}_G(t)$ and $\boldsymbol{\Sigma}_X(t)$ are diagonal matrices. It will be convenient to define $\tilde{\mathbf{Q}}(t, \tau)$, for $\tau \geq 0$, according to

$$\tilde{\mathbf{Q}}(t, \tau) = \mathbf{V}(\tau)^{-\top} \mathbf{Q}(t) \mathbf{V}(\tau)^{-1} \quad (117)$$

where $\mathbf{V}(\tau)$ is the same as defined in (116).

The remainder of the proof is split into two parts. In the first part we will prove that each eigenvalue of $\tilde{\mathbf{Q}}(t, \tau)$ has at most a single negative-to-nonnegative zero crossing. In the second part, we will show that this property

transfers to $\mathbf{Q}(t)$, thus completing the proof. Coincidentally, both parts of the proof will be based on contradiction arguments, i.e., we assume that the opposite of what we want to prove is true and, then, end up with an inconsistency.

1) *Single crossing point for the eigenvalues of $\tilde{\mathbf{Q}}(t, \tau)$* : Let us start by presenting a result on the differentiability of the eigenvalues of a symmetric matrix with respect to some scalar parameter t , which was studied by Rellich in [23, Ch. 1]¹:

Lemma 8: [23, Th. in p. 57] Suppose that $\mathbf{A}(t)$ is an n -dimensional symmetric matrix defined on some open interval $t \in (t_1, t_2)$. Suppose that the derivative $D_t \mathbf{A}(t)$ exists and it is continuous for each $t \in (t_1, t_2)$. Then, there exist n functions $\lambda_i(t)$, $i = 1, \dots, n$ with continuous derivatives in $t \in (t_1, t_2)$, such that

$$\mathbf{A}(t)\mathbf{u}_i(t) = \lambda_i(t)\mathbf{u}_i(t), \quad i = 1, \dots, n \quad (118)$$

for some properly chosen orthonormal system of vectors $\mathbf{u}_i(t)$, $i = 1, \dots, n$.

Since $\tilde{\mathbf{Q}}(t, \tau)$ is a symmetric matrix whose derivative $D_t \tilde{\mathbf{Q}}(t, \tau)$ exists, Lemma 8 ensures the existence of n continuous and differentiable functions such that they are equal to the eigenvalues of the matrix $\tilde{\mathbf{Q}}(t, \tau)$, for any choice of τ . These functions will be denoted from now on by $\lambda_i(t, \tau)$, for $i = 1, \dots, n$.

Now, let us assume that, at $t = t_0$, k of these eigenvalues (with $k \leq n$) are equal to zero, i.e., $\lambda_i(t_0, \tau) = 0$, for $i = 1, \dots, k$. Furthermore, we also assume that, from these k eigenvalues that are zero at $t = t_0$, s of them (with $s \leq k$) have a nonnegative-to-negative zero crossing at $t = t_0$. To sum up, we assume that the differentiable functions $\lambda_i(t, \tau)$, with $i = 1, \dots, s$ have a nonnegative-to-negative zero crossing at $t = t_0$.

Let us now present a property of differentiable functions that contain nonnegative-to-negative zero crossings:

Lemma 9: Assume that $f(t)$ has a nonnegative-to-negative zero crossing at $t = t_0$ and that $f(t)$ is differentiable in a neighborhood of t_0 . Then, there exists a positive value ε such that

$$f(t) < 0, \quad t \in (t_0, t_0 + \varepsilon), \quad (119)$$

$$D_t f(t) < 0, \quad t \in (t_0, t_0 + \varepsilon). \quad (120)$$

Proof: From Definition 2, (119) follows immediately for any $\varepsilon \leq \epsilon$. The proof for (120) follows easily from the mean value theorem and elementary calculus. ■

Applying Lemma 9 to the set of functions $\lambda_i(t, \tau)$, with $i = 1, \dots, s$, we readily obtain:

$$\left. \begin{aligned} \lambda_i(t, \tau) &< 0, & t &\in (t_0, t_0 + \varepsilon_i(\tau)) \\ D_t \lambda_i(t, \tau) &< 0, & t &\in (t_0, t_0 + \varepsilon_i(\tau)) \end{aligned} \right\} \quad i = 1, \dots, s \quad (121)$$

where we have written $\varepsilon_i(\tau)$ to make explicit the dependence of ε_i on the specific value of τ . For the sake of convenience, we want to eliminate the dependence of ε_i on τ . A possible method to eliminate this dependence is to define

$$\varepsilon_i^* = \inf_{\tau \in [t_0, t_0 + M]} \varepsilon_i(\tau) = \min_{\tau \in [t_0, t_0 + M]} \varepsilon_i(\tau) > 0 \quad (122)$$

¹Rellich studied the eigenvalue differentiability for Hermitian matrices. We specialized his result for the real case studied in this paper.

where, for the sake of convenience, we have restricted the values of τ in the interval $[t_0, t_0 + M]$, with M being an arbitrary fixed positive value (observe that since $\varepsilon_i(\tau)$ can be made arbitrarily small we can always guarantee that $M > \varepsilon_i(\tau) \geq \varepsilon_i^*$), and where the second equality follows from the fact that the optimization set is a closed interval and the third one follows from $\varepsilon_i(\tau) > 0, \forall \tau$.

Consequently, after this simplification, we have that, assuming that the differentiable functions $\lambda_i(t, \tau)$, with $i = 1, \dots, s$ have a nonnegative-to-negative zero crossing at $t = t_0$, they must fulfill:

$$\left. \begin{aligned} \lambda_i(t, \tau) &< 0, & t \in (t_0, t_0 + \varepsilon_i^*) \\ D_t \lambda_i(t, \tau) &< 0, & t \in (t_0, t_0 + \varepsilon_i^*) \end{aligned} \right\} \quad i = 1, \dots, s. \quad (123)$$

Now, we can particularize the expression above for the case where $t = t_0 + \varepsilon_i^*/2 \triangleq t^*$ and where we also choose $\tau = t^*$. We obtain

$$\left. \begin{aligned} \lambda_i(t^*, t^*) &< 0 \\ D_t \lambda_i(t, t^*)|_{t=t^*} &< 0 \end{aligned} \right\} \quad i = 1, \dots, s. \quad (124)$$

From this point our goal is to prove that the two conditions in (124) cannot both hold at the same time. For that purpose, we need an expression for the derivative of the eigenvalue function $D_t \lambda_i(t, t^*)$. Since we have that $\lambda_i(t_0, \tau) = 0$ for $i = 1, \dots, k$ (i.e., the multiplicity of the zero eigenvalue is k) we cannot guarantee that the multiplicity of the eigenvalue $\lambda_i(t^*, t^*)$ is equal to one. From this point, we assume that the multiplicity of $\lambda_i(t^*, t^*)$ is l .

Consequently, we now require the following result by Lancaster in [24, Th. 7] (it is also reproduced in [18, Ch. 8, Sec. 12, Th. 13]), which gives us an expression for the derivatives of the multiple eigenvalues²:

Lemma 10: [24, Th. 7] Under the assumptions in Lemma 8, let's consider the case where $\mathbf{A}(t)$ has a repeated eigenvalue λ_0 with multiplicity l , i.e., $\lambda_1(t) = \lambda_2(t) = \dots = \lambda_l(t) = \lambda_0$. Assume further that the $n \times l$ matrix $\mathbf{U}(t)$ spans the space associated with the repeated eigenvalues (i.e., $\mathbf{U}(t)$ contains one particular set of eigenvectors associated with the l repeated eigenvalue). Then, the l derivatives of the eigenvalues, which coincide at λ_0 are the eigenvalues of the matrix

$$\mathbf{U}(t)^T D_t \mathbf{A}(t) \mathbf{U}(t). \quad (125)$$

Using Lemma 10, we can write

$$D_t \lambda_i(t, \tau = t^*)|_{t=t^*} = \mu_i \left(\mathbf{1}_{n,l}^T D_t \tilde{\mathbf{Q}}(t, \tau = t^*) \Big|_{t=t^*} \mathbf{1}_{n,l} \right) \quad (126)$$

$$= \mu_i \left(\mathbf{1}_{n,l}^T \mathbf{V}(t^*)^{-T} D_t \mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t) \Big|_{t=t^*} \mathbf{V}(t^*)^{-1} \mathbf{1}_{n,l} \right) \quad (127)$$

$$\geq \mu_i \left([\boldsymbol{\Sigma}_{\mathbf{X}}(t^*) \mathbf{C}(t^*) \boldsymbol{\Sigma}_{\mathbf{X}}(t^*) - \boldsymbol{\Sigma}_G(t^*) \mathbf{C}(t^*) \boldsymbol{\Sigma}_G(t^*)]_{1:l, 1:l} \right) \quad (128)$$

$$= \mu_i \left([\boldsymbol{\Sigma}_{\mathbf{X}}(t^*)]_{1:l, 1:l} [\mathbf{C}(t^*)]_{1:l, 1:l} [\boldsymbol{\Sigma}_{\mathbf{X}}(t^*)]_{1:l, 1:l} - [\boldsymbol{\Sigma}_G(t^*)]_{1:l, 1:l} [\mathbf{C}(t^*)]_{1:l, 1:l} [\boldsymbol{\Sigma}_G(t^*)]_{1:l, 1:l} \right) \quad (129)$$

²The assumptions [24, Th. 7] are different than those in Lemma 8, but, once existence of the derivatives of the eigenvalues has been established, their expression has to be the same.

where $\mu_i(\mathbf{A})$ denotes the eigenvalue function of a generic matrix \mathbf{A} . Observe that, thanks to the fact that $\tilde{\mathbf{Q}}(t^*, t^*) = \Sigma_G(t^*) - \Sigma_X(t^*)$ is a diagonal matrix, in (126) we have chosen

$$\mathbf{U} = \mathbf{1}_{n,l} \triangleq \begin{pmatrix} \mathbf{I}_l \\ \mathbf{0}_{n-l,l} \end{pmatrix} \quad (130)$$

with \mathbf{I}_l being the $l \times l$ identity matrix and $\mathbf{0}_{n-l,l}$ being the $(n-l) \times l$ zero matrix. Moreover, in (128), we have used the fact that $\mathbf{A} \succeq \mathbf{B}$ implies both that $\mathbf{C}^\top \mathbf{A} \mathbf{C} \succeq \mathbf{C}^\top \mathbf{B} \mathbf{C}$ and that $\mu_i(\mathbf{A}) \geq \mu_i(\mathbf{B})$ [22, Cor. 7.7.4(c)] and the lower bound on the derivative of the matrix $\mathbf{Q}(t)$ given in Lemma 6. We further used the definition:

$$\mathbf{C}(t^*) = \mathbf{V}(t^*) \mathbf{B}(t^*) \mathbf{V}(t^*)^\top. \quad (131)$$

Observe that, since $\mathbf{B}(t^*)$ is a positive semidefinite diagonal matrix (see Lemma 4), we have $\mathbf{C}(t^*) \succeq \mathbf{0}$, which further implies that $[\mathbf{C}(t^*)]_{ii} \geq 0$, for all i . Finally, the upper-left $l \times l$ sub-matrix of matrix \mathbf{A} has been denoted by $[\mathbf{A}]_{1:l,1:l}$, and the last transition in (129) is due to the fact that both $\Sigma_X(t^*)$ and $\Sigma_G(t^*)$ are diagonal matrices.

In order to proceed with the proof, we require the following lemma.

Lemma 11: Let's consider a positive semidefinite matrix \mathbf{A} and two diagonal positive semidefinite matrices \mathbf{D}_1 and \mathbf{D}_2 such that $\mathbf{D}_1 \succeq \mathbf{D}_2 \succeq \mathbf{0}$. Then, we have that

$$\mu_{\max}(\mathbf{D}_1 \mathbf{A} \mathbf{D}_1 - \mathbf{D}_2 \mathbf{A} \mathbf{D}_2) \geq 0 \quad (132)$$

where μ_{\max} denotes the maximum eigenvalue function.

Proof: See Appendix A7. ■

Now, using the fact that $\mathbf{C}(t^*)$ is positive semidefinite, and the first condition in (124) that $\lambda_i(t^*, t^*) < 0$ for $i = 1, \dots, l$, which further implies that $[\Sigma_X(t^*)]_{1:l,1:l} \succ [\Sigma_G(t^*)]_{1:l,1:l} \succeq \mathbf{0}$ we can use Lemma 11 to conclude that,

$$\mu_{\max} \left([\Sigma_X(t^*)]_{1:l,1:l} [\mathbf{C}(t^*)]_{1:l,1:l} [\Sigma_X(t^*)]_{1:l,1:l} - [\Sigma_G(t^*)]_{1:l,1:l} [\mathbf{C}(t^*)]_{1:l,1:l} [\Sigma_G(t^*)]_{1:l,1:l} \right) \geq 0. \quad (133)$$

Last result together with (126)-(129) implies that there exists some $i \in [1, l]$ such that $\lambda_i(t^*, t^*) < 0$ and $D_t \lambda_i(t, \tau = t^*)|_{t=t^*} \geq 0$, which clearly contradicts the conditions in (124).

Since the contradiction described above holds for any arbitrary values for k , s , and l (under the condition $l \leq s \leq k \leq n$), we have thus proved that no nonnegative-to-negative zero crossing can occur for the eigenvalues of $\tilde{\mathbf{Q}}(t, \tau)$ or, equivalently, we have proved that the eigenvalues of $\tilde{\mathbf{Q}}(t, \tau)$ have at most a single negative-to-nonnegative zero crossing of the horizontal axis.

2) *Single crossing point for the eigenvalues of $\mathbf{Q}(t)$:* The relation between the sign of the eigenvalues of $\mathbf{Q}(t)$ and those of $\tilde{\mathbf{Q}}(t, \tau)$ is stated in the following lemma.

Lemma 12: For all τ and as a function of t , the number of positive, zero, and negative eigenvalues of $\mathbf{Q}(t)$ and $\tilde{\mathbf{Q}}(t, \tau)$ coincide.

Proof: The proof follows straightforwardly from the definition of $\tilde{\mathbf{Q}}(t, \tau)$, given in equation (117), and Sylvester's law of inertia for congruent matrices [25, p. 5]. ■

In the first part of the proof we have shown that $\tilde{\mathbf{Q}}(t, \tau)$ has, for each eigenvalue, at most, a single negative-to-nonnegative zero crossing. From this and Lemma 12, we can conclude that the number of negative eigenvalues of both functions cannot increase. Now, let's assume that $\mathbf{Q}(t)$ has an eigenvalue of multiplicity s with a nonnegative-to-negative zero crossing at t_0 , i.e., $\mu_i(\mathbf{Q}(t_0)) = 0$ and $\mu_i(\mathbf{Q}(t)) < 0$ for $t \in (t_0, t_0 + \varepsilon)$, for some positive ε and for $i = 1, \dots, s$. In order to refrain from increasing the number of negative eigenvalues, s negative eigenvalues at t_0 must become zero. However, if we examine the number of eigenvalues at $t_0 + \Delta$ for a sufficiently small Δ , the eigenvalues that were negative at t_0 are still negative at $t_0 + \Delta$, and the total number of negative eigenvalues has increased. Thus, contradicting the possibility of a nonnegative-to-negative zero crossing of the multiplicity s eigenvalue of $\mathbf{Q}(t)$. This is valid for any arbitrary t_0 , thus concluding our proof. ■

The following corollary is a simple consequence from Theorem 6.

Corollary 6: If for a given t' the function $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t') \succeq \mathbf{0}$ then for all $t \geq t'$ the function $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t) \succeq \mathbf{0}$.

B. The Conditioned Case

The results of the previous section can be simply extended to the conditioned case. Given an extension of the lower bound on the derivative of \mathbf{Q} , the extension of all other results is trivial. Thus, we briefly give the extension to the lower bound with a full proof (given in Appendix A8) and then for completeness restate the main result of this paper, for the conditioned case, without detailing the proof, which follows identically to the proof given above.

Lemma 13: The following lower bound holds:

$$\mathbf{D}_t \mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t) \succeq 2 \left(\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t) \mathbf{B}(t) \mathbf{E}_{\mathbf{X}|\mathbf{U}}^T(t) - \mathbf{E}_G(t) \mathbf{B}(t) \mathbf{E}_G^T(t) \right) \quad (134)$$

where $\mathbf{B}(t)$ was defined in (45), and assumed a positive semidefinite diagonal matrix for all t (see Lemma 4).

Proof: See Appendix A8. ■

Thus, the following theorem follows:

Theorem 7: Each eigenvalue of $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t)$ has, at most, a single negative-to-nonnegative zero crossing of the horizontal axis.

Proof: The proof follows the same steps as those in the proof of Theorem 6. ■

C. Properties of the Mutual Information

So far we have seen the “single crossing point” property of the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$, or more precisely, of its eigenvalues. As seen, this property also extends naturally to the conditioned case. In this section our goal is to relate this result to the mutual information between the input and the output of a parallel Gaussian channel. As expected, the advantage of this result is in the comparison between the mutual information assuming that the input to the channel has an arbitrary distribution and the mutual information assuming that it has a Gaussian distribution with an arbitrary covariance, $\mathbf{R}_{\mathbf{X}_G}$. Our goal is to make use of this result through the I-MMSE relationship, as given in equations (43)-(44) and (46)-(47). The results given in this section can be viewed as supporting theorem/lemmas, that make our “single crossing point” property applicable through the use of the I-MMSE relationship.

For clarity we will write the results in this section only for, the more general, conditioned case, from which one can easily derive the respective unconditioned theorems.

According to equation (47) the difference between the mutual information assuming that the input to the channel has an arbitrary distribution and the mutual information assuming that it has a Gaussian distribution with an arbitrary covariance, \mathbf{R}_{X_G} , is

$$\begin{aligned} I(\mathbf{X}_G; \mathbf{Y}(t)) - I(\mathbf{X}; \mathbf{Y}(t)|U) &= \int_{\tau=0}^t \text{Tr}(\mathbf{B}(\tau)(\mathbf{E}_G(\tau) - \mathbf{E}_{X|U}(\tau))) d\tau \\ &= \int_{\tau=0}^t \text{Tr}(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, \tau)) d\tau. \end{aligned} \quad (135)$$

Thus, we are interested in the properties of

$$\text{Tr}(\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)) = \sum_{i=1}^n \lambda_i(\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)) \quad (136)$$

where we have used the fact that the trace of a matrix \mathbf{A} is the sum of its eigenvalues [22, Th. 1.2.12]. The following theorem extends the “single crossing point” property of the eigenvalues of $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ to the eigenvalues of $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$.

Theorem 8: Each eigenvalue of $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ has, *at most*, a single negative-to-nonnegative zero crossing of the horizontal axis. Moreover, the eigenvalues of $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ have the following property:

$$\text{sign}\{\lambda_i(\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t))\} \in \{0, \text{sign}\{\lambda_i(\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t))\}\}. \quad (137)$$

Proof: For a non-singular $\mathbf{B}(t)$ and due to similarity [22, Cor. 1.3.4] we can write the following,

$$\lambda_i(\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)) = \lambda_i(\mathbf{B}^{\frac{1}{2}}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)\mathbf{B}^{\frac{1}{2}}(t)). \quad (138)$$

Recalling that $\mathbf{B}(t)$ is a positive semidefinite diagonal matrix, we have an eigenvalue of a congruent transformation. Thus, the proof follows similarly to the second part of the proof of Theorem 6 (given in Section V-A2), concluding the preservation of the signs of the eigenvalues of $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ in $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ and, as a result, concluding that all eigenvalues have, *at most* a single, negative-to-nonnegative zero crossing of the horizontal axis.

If $\mathbf{B}(t)$ is singular, we can assume without loss of generality that the i^{th} diagonal element is zero. Due to that, the i^{th} row of $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ is all zeros, that is, one of the eigenvalues of $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ is zero (and its sign is also zero). The rest of the eigenvalues can be calculated from the reduced problem, the matrix $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ without the i^{th} row and column. Recalling that $\mathbf{B}(t)$ is a diagonal matrix, this is simply the product of $\mathbf{B}(t)$ and $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ both without the i^{th} row and column. This procedure can be repeated as long as the reduced $\mathbf{B}(t)$ matrix is singular. When the reduced matrix is non-singular, we again follow the proof of Theorem 6.

Thus, we have shown that the eigenvalues preserve the sign of the eigenvalues of $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ with the additional possibility of falling to zero when $\mathbf{B}(t)$ becomes singular. ■

The next two lemmas provide the link between the above results, regarding the behavior of the eigenvalues of the matrix $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$ and the matrix $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{X_G}, t)$, and the mutual information. Thus, they facilitate

the usage of these results on information theory problems, as will be shown in the sequel. More particularly, so far we discussed the behavior of each and every eigenvalue of the matrix $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t)$ and the matrix $\mathbf{B}(t)\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t)$, which holds true for any proper choice of $\mathbf{R}_{\mathbf{X}_G}$ with no regards to the random vector \mathbf{X} . The next two lemmas identify the existence of specific Gaussian inputs which have unique properties with respect to the given random vector \mathbf{X} .

Lemma 14: Assume $\mathbf{X} \in \mathbb{R}^n$ is an arbitrary distributed random vector. For any $t_e \in [0, \infty)$, there exists a Gaussian input covariance matrix $\mathbf{R}_{\mathbf{X}_G}$ such that the following hold

- 1) $\mathbf{R}_{\mathbf{X}_G} \preceq \mathbf{R}_{\mathbf{X}}$
- 2) $I(\mathbf{X}; \mathbf{Y}(t_e)|\mathbf{U}) = I(\mathbf{X}_G; \mathbf{Y}_G(t_e))$
- 3) $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e) \succeq \mathbf{0}$

Proof: See Appendix A9. ■

Note that the above claim can be extended to a general non-singular $\mathbf{H}(t_e)$, that is, *not necessarily diagonal*, by defining $\widetilde{\mathbf{X}} \equiv \mathbf{H}(t_e)\mathbf{X}$. Due to the non-singularity of $\mathbf{H}(t_e)$, the mutual information is unchanged, i.e., $I(\widetilde{\mathbf{X}}; \mathbf{Y}(t_e)|\mathbf{U}) = I(\mathbf{X}; \mathbf{Y}(t_e)|\mathbf{U})$. Requirements 1 and 3 are preserved under any congruent transformation, specifically under the transformation $\mathbf{H}^{-1}(t_e)$.

The next lemma is an extension of Lemma 14 that will prove useful in the sequel.

Lemma 15: Assume that for a given input distribution on the pair (\mathbf{U}, \mathbf{X}) there exists a Gaussian random vector, \mathbf{X}_G^{ub} , with covariance $\mathbf{R}_{\mathbf{X}_G}^{\text{ub}}$ such that for some $t_e \in [0, \infty)$ we have that,

- 1) $I(\mathbf{X}; \mathbf{Y}(t_e)|\mathbf{U}) \leq I(\mathbf{X}_G^{\text{ub}}; \mathbf{Y}_G^{\text{ub}}(t_e))$
- 2) $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}^{\text{ub}}, t_e) \succeq \mathbf{0}$

Thus, there exists a Gaussian random vector, \mathbf{X}_G , with covariance $\mathbf{R}_{\mathbf{X}_G}$ such that the following holds:

- 1) $\mathbf{R}_{\mathbf{X}_G} \preceq \mathbf{R}_{\mathbf{X}_G}^{\text{ub}}$
- 2) $I(\mathbf{X}; \mathbf{Y}(t_e)|\mathbf{U}) = I(\mathbf{X}_G; \mathbf{Y}_G(t_e))$
- 3) $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e) \succeq \mathbf{0}$

Proof: The proof follows the proof of Lemma 14, where instead of using $\mathbf{E}_{\mathbf{X}}^{\text{lin}}(t_e)$ (209) as a trivial upper bound we use:

$$\mathbf{E}_G^{\text{ub}}(t_e) = \mathbf{I} - (\mathbf{R}_{\mathbf{X}_G}^{\text{ub}} + \mathbf{I})^{-1} \quad (139)$$

and the assumptions stated above. ■

D. Connections to Fisher Information

In addition to the MMSE matrix, another important quantity in estimation theory is the Fisher information matrix [26]. Its connection to information theory has been established in the late 1950's and has been attributed to de Bruijn [8]. The de Bruijn identity relates the derivative of the differential entropy to the Fisher information matrix

defined as³:

$$\mathbf{J}(\mathbf{Y}) = \mathbb{E}\{[\nabla \log P_{\mathbf{Y}}(\mathbf{Y})][\nabla \log P_{\mathbf{Y}}(\mathbf{Y})]^\top\} \quad (140)$$

where the expectation is over \mathbf{Y} . Note that this is a special form of the Fisher Information matrix (with respect to a translation parameter) which does not involve an explicit parameter as in its most general definition [26]. In [7] the authors have shown that the de Bruijn identity is equivalent to the I-MMSE relationship. Using this connection, the de Bruijn identity has been extended to a multivariate version in [17, Th. 4]. For our purposes we will use the following notation:

$$\mathbf{J}_{\mathbf{X}}(\mathbf{H}) = \mathbf{J}(\mathbf{H}\mathbf{X} + \mathbf{N}) \quad (141)$$

when we have some arbitrary input distribution on the random vector \mathbf{X} . For the case of a Gaussian distribution on \mathbf{X} with covariance matrix $\mathbf{R}_{\mathbf{X}_G}$ we will write $\mathbf{J}_G(\mathbf{R}_{\mathbf{X}_G}, \mathbf{H})$. We further note that, as in the case of the MMSE matrix, whenever the channel coefficients depend on other parameters, $\mathbf{H} = \mathbf{H}(\phi)$, we will write $\mathbf{J}_{\mathbf{X}}(\phi)$. We can now extend the idea of the the matrix \mathbf{Q} to the Fisher Information, using the following definition:

$$\mathbf{W}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, \phi) = \mathbf{J}_{\mathbf{X}}(\phi) - \mathbf{J}_G(\mathbf{R}_{\mathbf{X}_G}, \phi). \quad (142)$$

As in the case of the matrix \mathbf{Q} , the matrix \mathbf{W} has some distinct properties. Using the relationship between the two matrices we can derive these properties directly from the results of the previous sections. We first require the following lemma, given by Palomar and Verdú in [17].

Lemma 16: [17, App. E] Assuming the Gaussian additive noise channel (5), the following connection between the Fisher Information matrix and the MMSE matrix holds:

$$\mathbf{J}_{\mathbf{Y}} = \mathbf{I}_n - \mathbf{H}\mathbf{E}_{\mathbf{X}}\mathbf{H}^\top \quad (143)$$

Proof: The result follows directly from equation (106) in [17] by setting $\tilde{\Sigma}_n$ equal to the identity matrix and recalling that the MMSE matrix in (106) is the MMSE matrix of $\mathbf{Z} = \mathbf{H}\mathbf{X}$, from which it follows that $\mathbf{E}_{\mathbf{Z}} = \mathbf{H}\mathbf{E}_{\mathbf{X}}\mathbf{H}^\top$. ■

We can now state the main result of this section.

Theorem 9: The matrix $\mathbf{W}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ is related to the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ as follows:

$$\mathbf{W}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t) = \mathbf{H}(t)\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)\mathbf{H}(t)^\top. \quad (144)$$

Moreover, the properties given in Sections IV and V for the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$, transfer to the matrix $\mathbf{W}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$.

Proof: Equation (144) is obtained through the use of Lemma 16. The properties given in Section IV regarding the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ transfer to the matrix $\mathbf{W}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$, due to the fact that $\mathbf{H}(t)$ is a diagonal positive semidefinite matrix for all t . The properties given in Section V regarding the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ transfer to the matrix $\mathbf{W}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$, since it is simply a congruent transformation of $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{\mathbf{X}_G}, t)$ (this was explained in detail in part two of the proof of Theorem 6). ■

³For any differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, its gradient at any \mathbf{y} is a column vector $\nabla f(\mathbf{y}) = [D_{y_1}f(\mathbf{y}), \dots, D_{y_n}f(\mathbf{y})]^\top$.

E. Application: The Degraded Parallel Gaussian BC Capacity Region under Covariance Constraint

In this section we show that the result of Section V-C can be used to provide a converse proof for the *degraded* parallel Gaussian BC capacity region under a covariance constraint. We consider the following model:

$$\begin{aligned} \mathbf{Y}_1[m] &= \mathbf{H}_1 \mathbf{X}[m] + \mathbf{N}_1[m] \\ \mathbf{Y}_2[m] &= \mathbf{H}_2 \mathbf{X}[m] + \mathbf{N}_2[m] \end{aligned} \quad (145)$$

where $\mathbf{N}_1[m]$ and $\mathbf{N}_2[m]$ are standard additive Gaussian noise vectors independent for different time indices m , and \mathbf{H}_1 and \mathbf{H}_2 are diagonal positive semidefinite matrices such that $\mathbf{H}_1 \preceq \mathbf{H}_2$. $\mathbf{X} \in \mathbb{R}^n$ is the random input vector, and it is assumed independent for different time indices m .

We consider a covariance constraint:

$$\mathbf{R}_{\mathbf{X}} \preceq \mathbf{S} \quad (146)$$

where \mathbf{S} is some positive definite matrix.

Since we have a *degraded* BC, we can use the single-letter expression as given in (100). As in Section IV-D, we will follow the proof given for the scalar Gaussian BC in [6], [20]. Using Lemma 4 we can construct a path such that:

$$\begin{aligned} \mathbf{H}(t_2) &= \mathbf{H}_2 \\ \mathbf{H}(t_1) &= \mathbf{H}_1 \\ \mathbf{H}(0) &= \mathbf{0} \end{aligned} \quad (147)$$

where $0 \leq t_1 \leq t_2$ and $\mathbf{H}(t)$ is diagonal for all $t \in [0, t_2]$.

Now, assume a pair (\mathbf{U}, \mathbf{X}) with covariance $\mathbf{R}_{\mathbf{X}}$ for \mathbf{X} . According to Lemma 14, there exists a Gaussian random vector with covariance $\mathbf{R}_{\mathbf{X}_G}$ such that the following properties hold:

- 1) $\mathbf{R}_{\mathbf{X}_G} \preceq \mathbf{R}_{\mathbf{X}}$.
- 2) $I(\mathbf{X}; \mathbf{Y}(t_1)|\mathbf{U}) = I(\mathbf{X}_G; \mathbf{Y}_G(t_1))$.
- 3) $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t) \succeq \mathbf{0}$ for all $t \geq t_1$.

Using the I-MMSE relationship (47) we can write,

$$I(\mathbf{X}_G; \mathbf{H}(t)\mathbf{X}_G + \mathbf{N}) - I(\mathbf{X}; \mathbf{H}(t)\mathbf{X} + \mathbf{N}|\mathbf{U}) = \int_{\tau=0}^t \text{Tr}(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, \tau)) d\tau \quad (148)$$

$$= \int_{\tau=0}^t \sum_{i=1}^n \lambda_i(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, \tau)) d\tau. \quad (149)$$

Using the above properties on (149) we have that for any $t' > t_1$,

$$\begin{aligned} I(\mathbf{X}_G; \mathbf{H}(t')\mathbf{X}_G + \mathbf{N}) - I(\mathbf{X}; \mathbf{H}(t')\mathbf{X} + \mathbf{N}|U) &= \int_{\tau=0}^{t_1} \text{Tr}(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, \tau)) d\tau \\ &+ \int_{\tau=t_1}^{t'} \text{Tr}(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, \tau)) d\tau \end{aligned} \quad (150)$$

$$= 0 + \int_{\tau=t_1}^{t'} \sum_{i=1}^n \lambda_i(\mathbf{B}(\tau)\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, \tau)) d\tau \geq 0 \quad (151)$$

where (151) follows from property 2, and the inequality follows from property 3 and Theorem 8.

Thus, we have shown the existence of a Gaussian random vector, \mathbf{X}_G , with covariance matrix $\mathbf{R}_{\mathbf{X}_G}$, with the following properties:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}(t_1)|U) &= I(\mathbf{X}_G; \mathbf{Y}_G(t_1)) \\ I(\mathbf{X}; \mathbf{Y}(t_2)|U) &\leq I(\mathbf{X}_G; \mathbf{Y}_G(t_2)) \\ \mathbf{R}_{\mathbf{X}_G} &\preceq \mathbf{R}_{\mathbf{X}} \end{aligned} \quad (152)$$

Using these properties on the single-letter expression (100) we obtain the following outer bound,

$$\begin{aligned} R_1 &\leq I(U; \mathbf{Y}_1) = I(\mathbf{X}; \mathbf{Y}_1) - I(\mathbf{X}; \mathbf{Y}_1|U) \\ &\leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{H}_1^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{R}_{\mathbf{X}_G} \mathbf{H}_1^T| = \frac{1}{2} \log \frac{|\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{H}_1^T|}{|\mathbf{I} + \mathbf{H}_1 \mathbf{R}_{\mathbf{X}_G} \mathbf{H}_1^T|} \end{aligned} \quad (153)$$

$$R_2 \leq I(\mathbf{X}; \mathbf{Y}_2|U) \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_2 \mathbf{R}_{\mathbf{X}_G} \mathbf{H}_2^T| \quad (154)$$

This outer bound is tight and the achievability is well-known using superposition coding. This approach can be extended to the M-user scenario as shown in Appendix C.

F. Application: The Compound Degraded Parallel Gaussian BC Capacity Region under Covariance Constraint

In this section we show that the results of Section V-A can also be used to provide a converse proof for the compound *degraded* parallel Gaussian BC capacity region under a covariance constraint. We consider the following model,

$$\mathbf{Y}_{i_j}^j[m] = \mathbf{H}_{i_j}^j \mathbf{X}[m] + \mathbf{N}_{i_j}^j[m], \quad j = 1, \dots, M, \quad i_j = 1, \dots, K_j \quad (155)$$

where $\mathbf{N}_{i_j}^j$, $j = 1, \dots, M, i_j = 1, \dots, K_j$ are standard additive Gaussian noise vectors independent for different time indices m , and $\mathbf{H}_{i_j}^j$, $j = 1, \dots, M, i_j = 1, \dots, K_j$ are diagonal positive definite matrices such that:

$$\mathbf{H}_{i_j}^j \preceq \mathbf{H}_{i_{j+1}}^{j+1} \quad \forall j = 1, \dots, M, i_j \in \{1, \dots, K_j\}, i_{j+1} \in \{1, \dots, K_{j+1}\}. \quad (156)$$

Since these matrices are diagonal, there exist matrices $\mathbf{H}_{(j+1)j}^*$ for $j = 1, \dots, M-1$ such that

$$\mathbf{H}_{i_j}^j \preceq \mathbf{H}_{(j+1)j}^* \preceq \mathbf{H}_{i_{j+1}}^{j+1} \quad \forall j = 1, \dots, M-1, i_j \in \{1, \dots, K_j\}, i_{j+1} \in \{1, \dots, K_{j+1}\}. \quad (157)$$

Note that the equivalence between conditions (156) and (157) is not true in general (for non-diagonal matrices), as explained in [27]. $\mathbf{X} \in \mathbb{R}^n$ is the random input vector, and it is assumed independent for different time indices m . We consider a covariance constraint:

$$\mathbf{R}_X \preceq \mathbf{S} \quad (158)$$

where \mathbf{S} is some positive definite matrix.

Before proceeding, we provide the following single-letter expression for the capacity region of this M user memoryless channel. This is a simple extension of [27, Lem. 4].

Lemma 17: Consider a memoryless compound BC with input \mathbf{X} , M outputs $\mathbf{Y}_{i_j}^j$, $j = 1, \dots, M$, $i_j = 1, \dots, K_j$, and auxiliary random outputs $\mathbf{Y}_{(j+1)j}^*$ with $j \in \{1, \dots, M-1\}$. All outputs are defined by their conditional probability functions: $P_{\mathbf{Y}_{i_j}^j|\mathbf{X}}$ and $P_{\mathbf{Y}_{(j+1)j}^*|\mathbf{X}}$. Furthermore, assume that these outputs are stochastically *degraded* such that there exists some distribution such that $\mathbf{X} - \mathbf{Y}_{i_M}^M - \mathbf{Y}_{M(M-1)}^* - \mathbf{Y}_{i_{M-1}}^{M-1} - \mathbf{Y}_{(M-1)(M-2)}^* - \dots - \mathbf{Y}_{i_2}^2 - \mathbf{Y}_{21}^* - \mathbf{Y}_{i_1}^1$ form a Markov chain for every choice of i_1, i_2, \dots, i_M . The capacity region of this channel is given by the union of the rate tuples satisfying

$$R_j \leq \min_{i_j=1, \dots, K_j} I(\mathbf{V}_j; \mathbf{Y}_{i_j}^j | \mathbf{V}_{j-1}) \quad (159)$$

where $\mathbf{V}_0 \equiv \emptyset$, $\mathbf{V}_M \equiv \mathbf{X}$ and the union is over all probability distributions satisfying

$$\mathbf{V}_0 - \mathbf{V}_1 - \dots - \mathbf{V}_{M-1} - \mathbf{V}_M - \mathbf{X} - \mathbf{Y}_{i_M}^M - \mathbf{Y}_{M(M-1)}^* - \mathbf{Y}_{i_{M-1}}^{M-1} - \mathbf{Y}_{(M-1)(M-2)}^* - \dots - \mathbf{Y}_{i_2}^2 - \mathbf{Y}_{21}^* - \mathbf{Y}_{i_1}^1 \quad (160)$$

Proof: See Appendix D. ■

Using Lemma 17 we prove the following theorem,

Theorem 10: The capacity region of the compound *degraded* parallel Gaussian BC (155), is given by the following expression:

$$\begin{aligned} R_M &\leq \min_{i_M=1, \dots, K_M} \frac{1}{2} \log \left| \mathbf{H}_{i_M}^M \mathbf{R}_{G_M} (\mathbf{H}_{i_M}^M)^T + \mathbf{I} \right| \\ R_j &\leq \min_{i_j=1, \dots, K_j} \frac{1}{2} \log \left| \frac{\mathbf{H}_{i_j}^j \sum_{l=j}^M \mathbf{R}_{G_l} (\mathbf{H}_{i_j}^j)^T + \mathbf{I}}{\mathbf{H}_{i_j}^j \sum_{l=j+1}^M \mathbf{R}_{G_l} (\mathbf{H}_{i_j}^j)^T + \mathbf{I}} \right|, \quad \forall j = 1, \dots, M-1 \end{aligned} \quad (161)$$

where \mathbf{R}_{G_j} are some positive semidefinite matrices such that $\mathbf{0} \preceq \sum_{l=1}^M \mathbf{R}_{G_l} \preceq \mathbf{S}$.

Proof: According to Lemma 4 (and the remark after this lemma) for any set of $\{i_1, i_2, \dots, i_M\}$ where $i_j \in K_j$ we can construct a diagonal path such that

$$\begin{aligned} \mathbf{H}(t_{i_j}) &= \mathbf{H}_{i_j}, \quad j = 1, \dots, M \\ \mathbf{H}(t_{(j+1)j}) &= \mathbf{H}_{(j+1)j}^*, \quad j = 1, \dots, M-1 \\ \mathbf{H}(t=0) &= \mathbf{0} \end{aligned} \quad (162)$$

with $0 \leq t_{i_1} \leq t_{21} \leq t_{i_2} \leq \dots \leq t_{i_j} \leq t_{(j+1)j} \leq t_{i_{j+1}} \leq \dots \leq t_{i_M}$. Now, let's examine a tuple of rates on the boundary of the capacity region: $(R_1^{opt}, R_2^{opt}, \dots, R_M^{opt})$. Assume that this tuple has been attained by the joint distribution $P_{V_1, \dots, V_{M-1}, X}$ on the tuple with covariance $\mathbf{R}_X \preceq \mathbf{S}$ as required by the constraint (158).

We begin by looking at the following partial Markov chain:

$$V_0 - V_1 - \dots - V_{M-1} - V_M - X - Y_{M(M-1)}^* - Y_{(M-1)(M-2)}^* - \dots - Y_{21}^*. \quad (163)$$

Now, assuming that $Y_{(j+1)j}^*$ are the outputs, we can use Lemma 22 which states that there exist M Gaussian inputs X_{G_j} , with covariance matrices $\mathbf{R}_{X_{G_j}}$ such that,

$$\int_0^{t_{(j+1)j}} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, \tau) \right) d\tau = 0 \quad (164)$$

$$\int_0^{t_{(j+2)j}} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, \tau) \right) d\tau \geq 0, \quad \forall j = 1, \dots, M-2 \quad (165)$$

and such that $\mathbf{0} \preceq \mathbf{R}_{X_{G_j}} \preceq \mathbf{R}_{X_{G_{j-1}}}$, for $j = 2, \dots, M-1$ and $\mathbf{0} \preceq \mathbf{R}_{X_{G_1}} \preceq \mathbf{S}$. Furthermore,

$$\mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, t_{(j+1)j}) \succeq \mathbf{0} \quad (166)$$

for all $j = 1, \dots, M-1$.

Using this result, and according to Corollary 6 we know that $\mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, t) \succeq \mathbf{0}$ for all $t \geq t_{(j+1)j}$. This holds for any diagonal path, such that $\mathbf{H}(t_{(j+1)j}) = \mathbf{H}_{(j+1)j}^*$. Now, using Theorem 8 and (164) we can conclude that,

$$\begin{aligned} \int_0^t \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, \tau) \right) d\tau &\leq 0, \quad \forall t \leq t_{(j+1)j} \\ \int_0^t \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, \tau) \right) d\tau &\geq 0, \quad \forall t \geq t_{(j+1)j}. \end{aligned} \quad (167)$$

Due to the Markov chain:

$$V_j - V_{j+1} - X - Y_{i_{j+1}}^{j+1} - Y_{(j+1)j}^* - Y_{i_j}^j \quad (168)$$

(167) is particularly valid for,

$$\begin{aligned} \int_0^{t_{i_j}} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, \tau) \right) d\tau &\leq 0, \quad \forall i_j \in K_j \\ \int_0^{t_{i_{j+1}}} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(X|V_j, \mathbf{R}_{X_{G_j}}, \tau) \right) d\tau &\geq 0, \quad \forall i_{j+1} \in K_{j+1} \end{aligned} \quad (169)$$

for any $j = 1, \dots, M-1$. Equations (159) can be written explicitly, as follows:

$$\begin{aligned}
R_M &\leq \min_{i_M=1, \dots, K_M} I(\mathbf{X}; \mathbf{Y}_{i_M}^M | \mathbf{V}_{M-1}) \\
R_{M-1} &\leq \min_{i_{M-1}=1, \dots, K_{M-1}} I(\mathbf{X}; \mathbf{Y}_{i_{M-1}}^{M-1} | \mathbf{V}_{M-2}) - I(\mathbf{X}; \mathbf{Y}_{i_{M-1}}^{M-1} | \mathbf{V}_{M-1}) \\
R_{M-2} &\leq \min_{i_{M-2}=1, \dots, K_{M-2}} I(\mathbf{X}; \mathbf{Y}_{i_{M-2}}^{M-2} | \mathbf{V}_{M-3}) - I(\mathbf{X}; \mathbf{Y}_{i_{M-2}}^{M-2} | \mathbf{V}_{M-2}) \\
&\vdots \\
R_2 &\leq \min_{i_2=1, \dots, K_2} I(\mathbf{X}; \mathbf{Y}_{i_2}^2 | \mathbf{V}_1) - I(\mathbf{X}; \mathbf{Y}_{i_2}^2 | \mathbf{V}_2) \\
R_1 &\leq \min_{i_1=1, \dots, K_1} I(\mathbf{X}; \mathbf{Y}_{i_1}^1 | \mathbf{V}_0 \equiv \emptyset) - I(\mathbf{X}; \mathbf{Y}_{i_1}^1 | \mathbf{V}_1).
\end{aligned} \tag{170}$$

Using (169) and the trivial bound on $I(\mathbf{X}; \mathbf{Y}_{i_1}^1)$ we can upper bound these expressions as follows:

$$\begin{aligned}
R_M &\leq \min_{i_M=1, \dots, K_M} I(\mathbf{X}_{G_{M-1}}; \mathbf{H}_{i_M}^M \mathbf{X}_{G_{M-1}} + \mathbf{N}) \\
R_{M-1} &\leq \min_{i_{M-1}=1, \dots, K_{M-1}} I(\mathbf{X}_{G_{M-2}}; \mathbf{H}_{i_{M-1}}^{M-1} \mathbf{X}_{G_{M-2}} + \mathbf{N}) - I(\mathbf{X}_{G_{M-1}}; \mathbf{H}_{i_{M-1}}^{M-1} \mathbf{X}_{G_{M-1}} + \mathbf{N}) \\
R_{M-2} &\leq \min_{i_{M-2}=1, \dots, K_{M-2}} I(\mathbf{X}_{G_{M-3}}; \mathbf{H}_{i_{M-2}}^{M-2} \mathbf{X}_{G_{M-3}} + \mathbf{N}) - I(\mathbf{X}_{G_{M-2}}; \mathbf{H}_{i_{M-2}}^{M-2} \mathbf{X}_{G_{M-2}} + \mathbf{N}) \\
&\vdots \\
R_2 &\leq \min_{i_2=1, \dots, K_2} I(\mathbf{X}_{G_1}; \mathbf{H}_{i_2}^2 \mathbf{X}_{G_1} + \mathbf{N}) - I(\mathbf{X}_{G_2}; \mathbf{H}_{i_2}^2 \mathbf{X}_{G_2} + \mathbf{N}) \\
R_1 &\leq \min_{i_1=1, \dots, K_1} \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_{i_1}^1 \mathbf{S} (\mathbf{H}_{i_1}^1)^T| - I(\mathbf{X}_{G_1}; \mathbf{H}_{i_1}^1 \mathbf{X}_{G_1} + \mathbf{N}).
\end{aligned} \tag{171}$$

Defining,

$$\begin{aligned}
\mathbf{R}_{G1} &= \mathbf{S} - \mathbf{R}_{\mathbf{X}_{G1}} \\
\mathbf{R}_{Gj} &= \mathbf{R}_{\mathbf{X}_{Gj-1}} - \mathbf{R}_{\mathbf{X}_{Gj}}, \quad \forall j = 2, \dots, M-1 \\
\mathbf{R}_{GM} &= \mathbf{R}_{\mathbf{X}_{GM-1}}
\end{aligned} \tag{172}$$

(171) becomes the following set of upper bound,

$$\begin{aligned}
R_M &\leq \min_{i_M=1, \dots, K_M} \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_{i_M}^M \mathbf{R}_{GM} (\mathbf{H}_{i_M}^M)^T| \\
R_j &\leq \min_{i_j=1, \dots, K_j} \frac{1}{2} \log \left| \frac{\mathbf{I} + \mathbf{H}_{i_j}^j \sum_{l=j}^M \mathbf{R}_{Gl} (\mathbf{H}_{i_j}^j)^T}{\mathbf{I} + \mathbf{H}_{i_j}^j \sum_{l=j+1}^M \mathbf{R}_{Gl} (\mathbf{H}_{i_j}^j)^T} \right|, \quad \forall j = 1, \dots, M-1
\end{aligned} \tag{173}$$

where \mathbf{R}_{Gj} are some positive semidefinite matrices such that $\mathbf{0} \preceq \sum_{l=1}^M \mathbf{R}_{Gl} = \mathbf{S}$.

The above upper bounds can be attained simultaneously using a joint Gaussian distribution on the tuple,

$$(\mathbf{V}_0 \equiv \emptyset, \mathbf{V}_1, \dots, \mathbf{V}_{M-1}, \mathbf{V}_M \equiv \mathbf{X}) \tag{174}$$

as follows:

$$\mathbf{V}_j = \mathbf{V}_{j-1} + \mathbf{U}_j \quad (175)$$

where $\mathbf{U}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{G_j})$ for $j = 1, \dots, M$, independent of each other, and where \mathbf{R}_{G_j} are positive semidefinite matrices such that $\sum_{l=1}^M \mathbf{R}_{G_l} \preceq \mathbf{S}$. This concludes the proof of the capacity region. ■

VI. SUMMARY

In this work we extended the “single crossing point” property from the scalar setting to the parallel MIMO setting. We have shown three different “single crossing point” properties, given in three phases of extension from scalar-to-vector. These properties cannot be trivially deduced from each other. All three emphasize the basic optimality of the Gaussian input distribution in the Gaussian regime. The most general of these properties, given in the third phase, shows a “single crossing point” property for each of the eigenvalues of the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{X_G}, t)$, the difference between the MMSE matrix assuming an arbitrary Gaussian input, and the MMSE matrix assuming an arbitrary input distribution. We demonstrate the applicability of these properties on several information theoretic problems: a proof of a special case of Shannon’s vector EPI, a converse proof of the capacity region of the parallel *degraded* MIMO broadcast channel (BC) under per-antenna power constraints and under covariance constraint, and a converse proof of the capacity region of the compound parallel *degraded* MIMO BC under covariance constraint.

An open question is: can we extend the “single crossing point” property to the general MIMO channel? Note that, although the optimality of the Gaussian input is known for several MIMO Gaussian multi-terminal problems, we cannot necessarily conclude the existence of a “single crossing point” property. However, the implications of a general “single crossing point” property go beyond the specific applications shown here, and are also of interest on their own.

APPENDIX

A. Proofs of Lemmas

1) *Proof of Lemma 1:* Since \mathbf{A} is positive semidefinite we can always write $\mathbf{A} = \alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^T$ such that $\text{Tr}(\bar{\mathbf{A}} \bar{\mathbf{A}}^T) = n$ and $\alpha \geq 0$. Then, it can be checked that

$$q_{\mathbf{A}}(\mathbf{X}, \sigma^2, \gamma) = \frac{\sigma^2}{1 + \sigma^2 \gamma} \text{Tr}(\mathbf{A}) - \text{Tr}(\mathbf{A} \mathbf{E}_{\mathbf{X}}(\gamma)) \quad (176)$$

$$= \alpha \left(n \frac{\sigma^2}{1 + \sigma^2 \gamma} - \text{Tr}(\bar{\mathbf{A}}^T \mathbf{E}_{\mathbf{X}}(\gamma) \bar{\mathbf{A}}) \right) \quad (177)$$

$$= \alpha \left(n \frac{\sigma^2}{1 + \sigma^2 \gamma} - \text{Tr}(\mathbf{E}_{\bar{\mathbf{A}}^T \mathbf{X}}(\gamma)) \right) \quad (178)$$

$$= \alpha q_{\mathbf{I}_n}(\hat{\mathbf{X}}, \sigma^2, \gamma) \quad (179)$$

where we have defined $\hat{\mathbf{X}} = \bar{\mathbf{A}}^T \mathbf{X}$. Now, from (179) and the fact that $\alpha \geq 0$, the desired result follows. ■

2) *Proof of Lemma 2:* Let us consider the random vector $\mathbf{X} \in \mathbb{R}^n$, whose covariance is given by \mathbf{R}_X and denote its eigenvalues by $\lambda_{X,i}$. Recalling the model in (13), it is well known that $\mathbf{E}_X(\gamma) \preceq \mathbf{R}_X - \gamma \mathbf{R}_X (\gamma \mathbf{R}_X + \mathbf{I}_n)^{-1} \mathbf{R}_X$, [26]. Thus, we have that

$$\text{Tr}(\mathbf{E}_X(\gamma)) \leq \text{Tr}(\mathbf{R}_X - \gamma \mathbf{R}_X (\gamma \mathbf{R}_X + \mathbf{I}_n)^{-1} \mathbf{R}_X) \quad (180)$$

$$= \sum_{i=1}^n \left(\lambda_{X,i} - \frac{\gamma \lambda_{X,i}^2}{1 + \gamma \lambda_{X,i}} \right) \quad (181)$$

$$= \sum_{i=1}^n \frac{\lambda_{X,i}}{1 + \gamma \lambda_{X,i}}. \quad (182)$$

Now, realizing that the right hand side in (182) is a Schur-concave function (it follows directly from the concavity of $\frac{\lambda}{1+\gamma\lambda}$) and that, from the statement of Lemma 2, we have that $\sum_{i=1}^n \lambda_{X,i} \leq n\sigma^2$, it follows directly from majorization theory [28] that the right hand side in (182) is maximized when $\lambda_{X,i}$ are uniformly distributed, i.e., $\lambda_{X,i} = \sigma^2$. ■

3) *Proof of Lemma 3:* From the definition in (16), it follows that

$$D_\gamma q_A(\mathbf{X}, \sigma^2, \gamma) = -\frac{\sigma^4}{(1 + \sigma^2 \gamma)^2} \text{Tr}(\mathbf{A}) - D_\gamma \text{Tr}(\mathbf{A} \mathbf{E}_X(\gamma)). \quad (183)$$

The expression for $D_\gamma \text{Tr}(\mathbf{A} \mathbf{E}_X(\gamma))$ can be computed from the results in [14] and applying the chain rule as

$$D_\gamma \text{Tr}(\mathbf{A} \mathbf{E}_X(\gamma)) = D_{\mathbf{E}_X(\gamma)} \text{Tr}(\mathbf{A} \mathbf{E}_X(\gamma)) \cdot D_{\mathbf{H}} \mathbf{E}_X(\gamma) \cdot D_\gamma \mathbf{H} \quad (184)$$

$$= \text{vec}^\top(\mathbf{A}^\top) \mathbf{D}_n (-2 \mathbf{D}_n^\top \mathbf{E}\{\Phi_X(\mathbf{Y}) \otimes \Phi_X(\mathbf{Y})\} (\mathbf{I}_n \otimes \mathbf{H}^\top)) \frac{1}{2\sqrt{\gamma}} \text{vec}(\mathbf{I}_n) \quad (185)$$

$$= -\text{vec}^\top(\mathbf{A}^\top) \mathbf{N}_n \mathbf{E}\{\Phi_X(\mathbf{Y}) \otimes \Phi_X(\mathbf{Y})\} \text{vec}(\mathbf{I}_n) \quad (186)$$

$$= -\text{Tr}(\mathbf{A} \mathbf{E}\{\Phi_X(\mathbf{Y})^2\}) \quad (187)$$

where we have used that $\mathbf{H} = \sqrt{\gamma} \mathbf{I}_n$, $\mathbf{N}_n \mathbf{E}\{\Phi_X(\mathbf{Y}) \otimes \Phi_X(\mathbf{Y})\} = \mathbf{E}\{\Phi_X(\mathbf{Y}) \otimes \Phi_X(\mathbf{Y})\} \mathbf{N}_n$, and $\mathbf{N}_n \text{vec}(\mathbf{I}_n) = \text{vec}(\mathbf{I}_n)$ (see [14, App. A] for the definitions of the matrices \mathbf{D}_n and \mathbf{N}_n and some of their properties).

Plugging (187) in (183), the desired result follows. ■

4) *Proof of Lemma 5:* For any arbitrarily distributed random vector \mathbf{X} , with zero mean (assumed w.l.o.g.) and covariance matrix given by \mathbf{R}_X , it is well known that $\mathbf{E}_X(t) \preceq \mathbf{E}_G(\mathbf{R}_X, t)$, from which it follows that [22, Obs. 7.1.2]

$$[\mathbf{E}_X(t)]_{ii} \leq [\mathbf{E}_G(\mathbf{R}_X, t)]_{ii} \quad (188)$$

where we recall that $\mathbf{E}_G(\mathbf{R}_X, t)$ is the MMSE matrix attained assuming a zero mean Gaussian input with covariance matrix equal to \mathbf{R}_X . Observe that equality in (188) is attained if and only if $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_X)$.

Furthermore, from the fact that dependence among entries can only improve the MMSE, we have:

$$[\mathbf{E}_G(\mathbf{R}_X, t)]_{ii} \leq [\mathbf{E}_G(\mathbf{I}_n \circ \mathbf{R}_X, t)]_{ii} = \frac{[\mathbf{R}_X]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2 [\mathbf{R}_X]_{ii}} \quad (189)$$

where $\mathbf{E}_G(\mathbf{I}_n \circ \mathbf{R}_X, t)$ represents the MMSE matrix when the entries of the input vector are independent Gaussian random variables (thus, with diagonal covariance matrix). Observe that equality in (189) is obtained if and only if the entries of the Gaussian distribution in the left hand side are independent.

Now, the desired result follows immediately from the fact that the right hand side in (189) is an increasing function of $[\mathbf{R}_X]_{ii}$. ■

5) *Proof of Lemma 6:* We first provide the derivative of the MMSE with respect to the parameter t . Using equation (52), we have

$$D_t [\mathbf{E}_X(t)]_{ij} = \sum_l D_{[\mathbf{H}(t)]_{ll}} [\mathbf{E}_X(t)]_{ij} [D_t \mathbf{H}(t)]_{ll}. \quad (190)$$

Using the result ([14, eq. (131)]),

$$\begin{aligned} D_{[\mathbf{H}(t)]_{ll}} [\mathbf{E}_X(t)]_{ij} &= -\mathbb{E}\{[\Phi_X(\mathbf{Y})]_{jl} [\Phi_X(\mathbf{Y})\mathbf{H}(t)^T]_{il} + [\Phi_X(\mathbf{Y})]_{il} [\Phi_X(\mathbf{Y})\mathbf{H}(t)^T]_{jl}\} \\ &= -\mathbb{E}\{[\Phi_X(\mathbf{Y})]_{jl} [\Phi_X(\mathbf{Y})]_{il} [\mathbf{H}(t)]_{ll} + [\Phi_X(\mathbf{Y})]_{il} [\Phi_X(\mathbf{Y})]_{jl} [\mathbf{H}(t)]_{ll}\} \\ &= -2[\mathbf{H}(t)]_{ll} \mathbb{E}\{[\Phi_X(\mathbf{Y})]_{jl} [\Phi_X(\mathbf{Y})]_{il}\} \end{aligned} \quad (191)$$

where $\Phi_X(\mathbf{y})$ was defined in (9). The second equality in equation (191) is due to the fact that $\mathbf{H}(t)$ is diagonal. Thus, we can write the derivative of $[\mathbf{E}_X(t)]_{ij}$ as

$$\begin{aligned} D_t [\mathbf{E}_X(t)]_{ij} &= -2 \sum_l [\mathbf{H}(t)]_{ll} \mathbb{E}\{[\Phi_X(\mathbf{Y})]_{jl} [\Phi_X(\mathbf{Y})]_{il}\} [D_t \mathbf{H}(t)]_{ll} \\ &= -2 \sum_l [\mathbf{B}(t)]_{ll} \mathbb{E}\{[\Phi_X(\mathbf{Y})]_{jl} [\Phi_X(\mathbf{Y})]_{il}\} \end{aligned} \quad (192)$$

since $[\mathbf{B}(t)]_{ll} = [\mathbf{H}(t)]_{ll} [D_t \mathbf{H}(t)]_{ll}$ (45). We can put this expression into a matrix form as follows:

$$D_t \mathbf{E}_X(t) = -2 \sum_l [\mathbf{B}(t)]_{ll} \mathbb{E}\{[\Phi_X(\mathbf{Y})]_l [\Phi_X(\mathbf{Y})]_l^T\} \quad (193)$$

where $[\Phi_X(\mathbf{y})]_l$ is the l^{th} column of the matrix $\Phi_X(\mathbf{y})$. Using the fact that for a Gaussian input distribution $\Phi_X(\mathbf{y})$ does not depend on \mathbf{Y} and thus $\Phi_X(\mathbf{y}) = \mathbb{E}\{\Phi_X(\mathbf{y})\} = \mathbf{E}_G(t)$ [14], we can obtain the following lower bound on the derivative of the matrix $\mathbf{Q}(\mathbf{X}, \mathbf{R}_{X_G}, t)$:

$$\begin{aligned} D_t \mathbf{Q}(\mathbf{X}, \mathbf{R}_{X_G}, t) &= 2 \sum_l [\mathbf{B}(t)]_{ll} \left(\mathbb{E}\{[\Phi_X(\mathbf{Y})]_l [\Phi_X(\mathbf{Y})]_l^T\} - [\mathbf{E}_G]_l [\mathbf{E}_G]_l^T \right) \\ &\succeq 2 \sum_l [\mathbf{B}(t)]_{ll} \left(\mathbb{E}\{[\Phi_X(\mathbf{Y})]_l\} \mathbb{E}\{[\Phi_X(\mathbf{Y})]_l\}^T - [\mathbf{E}_G]_l [\mathbf{E}_G]_l^T \right) \\ &= 2 \sum_l [\mathbf{B}(t)]_{ll} \left(\mathbf{E}_{X_l} \mathbf{E}_{X_l}^T - [\mathbf{E}_G]_l [\mathbf{E}_G]_l^T \right) \\ &= 2 (\mathbf{E}_X(t) \mathbf{B}(t) \mathbf{E}_X^T(t) - \mathbf{E}_G(t) \mathbf{B}(t) \mathbf{E}_G^T(t)) \end{aligned}$$

where the inequality is due to Jensen. This concludes the proof of the lemma. ■

6) *Proof of Lemma 7:* Since \mathbf{A} and \mathbf{B} are two general positive semidefinite matrices, the dimension of the intersection of their null spaces, denoted by $N(\cdot)$ fulfills

$$\dim N(\mathbf{A}) \cap N(\mathbf{B}) = k, \quad 0 \leq k \leq n. \quad (194)$$

Let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be an orthonormal basis of the n -dimensional space such that $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an orthonormal basis of $N(\mathbf{A}) \cap N(\mathbf{B})$ and define $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_n]$. We thus have

$$\mathbf{U}^\top \mathbf{A} \mathbf{U} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}' \end{pmatrix}, \quad \mathbf{U}^\top \mathbf{B} \mathbf{U} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}' \end{pmatrix}, \quad (195)$$

where \mathbf{A}' and \mathbf{B}' are the non-zero $(n - k) \times (n - k)$ lower right square sub-matrices of $\mathbf{U}^\top \mathbf{A} \mathbf{U}$ and $\mathbf{U}^\top \mathbf{B} \mathbf{U}$, respectively. Observe that now we have $N(\mathbf{A}') \cap N(\mathbf{B}') = \{\emptyset\}$.

Now, from [22, Sec. 4.5, Prob. 8(e)], we have that \mathbf{A} and \mathbf{B} are simultaneously diagonalizable by an invertible matrix \mathbf{S} if and only if \mathbf{A}' and \mathbf{B}' are also simultaneously diagonalizable. Consequently, we have reduced our proof to showing the simultaneous diagonalization of two positive semidefinite matrices such that the dimension of the intersection of their null spaces is 0.

From this point, we can thus assume the following:

$$\mathbf{A} = \tilde{\mathbf{A}}^\top \tilde{\mathbf{A}} \succeq \mathbf{0}, \quad (196)$$

$$\mathbf{B} = \tilde{\mathbf{B}}^\top \tilde{\mathbf{B}} \succeq \mathbf{0}, \quad (197)$$

$$\dim N(\mathbf{A}) \cap N(\mathbf{B}) = 0. \quad (198)$$

The next step is to prove that \mathbf{A} and \mathbf{B} have no common isotropic vector, which is defined in [29, Def. 1.7.14] as a vector $\mathbf{x} \neq \mathbf{0}$ such that $\mathbf{x}^\top \mathbf{A} \mathbf{x} = 0$ and $\mathbf{x}^\top \mathbf{B} \mathbf{x} = 0$ are both simultaneously fulfilled.

Using the expression in (196), we have that

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = 0 \Leftrightarrow \mathbf{x}^\top \tilde{\mathbf{A}}^\top \tilde{\mathbf{A}} \mathbf{x} = 0 \Leftrightarrow \tilde{\mathbf{A}} \mathbf{x} = \mathbf{0}, \quad (199)$$

which can also be applied to $\mathbf{x}^\top \mathbf{B} \mathbf{x} = 0$. Consequently, if a vector \mathbf{x} fulfills $\mathbf{x}^\top \mathbf{A} \mathbf{x} = 0$ and $\mathbf{x}^\top \mathbf{B} \mathbf{x} = 0$, we have necessarily that $\mathbf{x} \in N(\tilde{\mathbf{A}}) \cap N(\tilde{\mathbf{B}})$. However, since $N(\tilde{\mathbf{A}}) \subseteq N(\mathbf{A})$ and, similarly, $N(\tilde{\mathbf{B}}) \subseteq N(\mathbf{B})$, from (198) we have that $\dim N(\tilde{\mathbf{A}}) \cap N(\tilde{\mathbf{B}}) = 0$, which implies that \mathbf{A} and \mathbf{B} have no common isotropic vector. Now, from [29, Th. 1.7.17] we have that \mathbf{A} and \mathbf{B} are simultaneously diagonalizable. ■

7) *Proof of Lemma 11:* For this proof we require the following result:

Lemma 18: Let's consider two positive semidefinite matrices \mathbf{A} and \mathbf{B} . Then we have

$$\mu_{\max}(\mathbf{A} - \mathbf{B}) \geq \mu_{\max}(\mathbf{A}) - \mu_{\max}(\mathbf{B}) \quad (200)$$

where we recall that $\mu_{\max}(\mathbf{A})$ denotes the maximum eigenvalue of matrix \mathbf{A} .

Proof: The proof follows directly from [22, Th. 4.3.1] recalling that $\mu_{\min}(-\mathbf{B}) = -\mu_{\max}(\mathbf{B})$. \blacksquare

Now, it is clear that for a positive semidefinite matrix \mathbf{A} and two positive semidefinite diagonal matrices \mathbf{D}_1 and \mathbf{D}_2 we have that $\mathbf{D}_i \mathbf{A} \mathbf{D}_i \succeq \mathbf{0}$, $i = 1, 2$. Now, from the above lemma we have,

$$\begin{aligned} \mu_{\max}(\mathbf{D}_1 \mathbf{A} \mathbf{D}_1 - \mathbf{D}_2 \mathbf{A} \mathbf{D}_2) &\geq \mu_{\max}(\mathbf{D}_1 \mathbf{A} \mathbf{D}_1) - \mu_{\max}(\mathbf{D}_2 \mathbf{A} \mathbf{D}_2) \\ &= \mu_{\max}(\mathbf{A}^{\frac{1}{2}} \mathbf{D}_1^2 \mathbf{A}^{\frac{1}{2}}) - \mu_{\max}(\mathbf{A}^{\frac{1}{2}} \mathbf{D}_2^2 \mathbf{A}^{\frac{1}{2}}) \end{aligned} \quad (201)$$

where the last equality follows from [22, Th. 1.3.20] and the remark, for square matrices, in the paragraph preceding it. Finally, since $\mathbf{D}_1 \succeq \mathbf{D}_2 \succeq \mathbf{0}$ and they are both diagonal, we have that $\mathbf{D}_1^2 \succeq \mathbf{D}_2^2 \succeq \mathbf{0}$ and, using [22, Obs. 7.7.2] and [22, Cor. 7.7.4] we can write,

$$\mu_{\max}(\mathbf{A}^{\frac{1}{2}} \mathbf{D}_1^2 \mathbf{A}^{\frac{1}{2}}) \geq \mu_{\max}(\mathbf{A}^{\frac{1}{2}} \mathbf{D}_2^2 \mathbf{A}^{\frac{1}{2}}) \quad (202)$$

from which the desired result follows. \blacksquare

8) *Proof of Lemma 13:* We extend the lower bound derived in Lemma 6 to the conditioned case, that is, we assume $\mathbf{U} - \mathbf{X} - \mathbf{Y}$. From (193), for the conditioned case we have the following:

$$\begin{aligned} \mathbf{D}_t \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t, \mathbf{u}) &= -2 \sum_l [\mathbf{B}(t)]_{ll} \mathbf{E}\{[\Phi_{\mathbf{X}_u}(\mathbf{Y})]_l [\Phi_{\mathbf{X}_u}(\mathbf{Y})]_l^T\} \\ &= -2 \sum_l [\mathbf{B}(t)]_{ll} \mathbf{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U} = \mathbf{u})]_l [\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U} = \mathbf{u})]_l^T\}. \end{aligned} \quad (203)$$

Taking expectation according to \mathbf{U} on both sides we have:

$$\mathbf{D}_t \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t) = \mathbf{E}\{\mathbf{D}_t \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t, \mathbf{U})\} = -2 \sum_l [\mathbf{B}(t)]_{ll} \mathbf{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l [\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l^T\}. \quad (204)$$

The derivative of $\mathbf{Q}(\mathbf{X}|\mathbf{Y}, \mathbf{R}_{\mathbf{X}_G}, t)$ is then given by:

$$\begin{aligned} \mathbf{D}_t \mathbf{Q}(\mathbf{X}|\mathbf{Y}, \mathbf{R}_{\mathbf{X}_G}, t) &= 2 \sum_l [\mathbf{B}(t)]_{ll} \left(\mathbf{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l [\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l^T\} - \mathbf{E}\{[\Phi_{\mathbf{X}_G}(\mathbf{Y})]_l [\Phi_{\mathbf{X}_G}(\mathbf{Y})]_l^T\} \right) \\ &= 2 \sum_l [\mathbf{B}(t)]_{ll} \left(\mathbf{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l [\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l^T\} - [\mathbf{E}_G(t)]_l [\mathbf{E}_G(t)]_l^T \right) \\ &\succeq 2 \sum_l [\mathbf{B}(t)]_{ll} \left(\mathbf{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l\} \mathbf{E}\{[\Phi_{\mathbf{X}}(\mathbf{Y}, \mathbf{U})]_l^T\} - [\mathbf{E}_G(t)]_l [\mathbf{E}_G(t)]_l^T \right) \\ &= 2 \sum_l [\mathbf{B}(t)]_{ll} \left([\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t)]_l [\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t)]_l^T - [\mathbf{E}_G(t)]_l [\mathbf{E}_G(t)]_l^T \right) \\ &= 2 \left(\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t) \mathbf{B}(t) \mathbf{E}_{\mathbf{X}|\mathbf{U}}^T(t) - \mathbf{E}_G(t) \mathbf{B}(t) \mathbf{E}_G^T(t) \right) \end{aligned} \quad (205)$$

where the inequality is due to Jensen. This completes the proof of the lemma. \blacksquare

9) *Proof of Lemma 14:* We first claim that w.l.o.g. we can restrict the proof to $\mathbf{H}(t_e) = \mathbf{I}$. This is shown by redefining $\widetilde{\mathbf{X}} = \mathbf{H}(t_e)\mathbf{X}$. Now, if $\mathbf{H}(t_e)$ is non-singular, then this redefinition does not change the mutual information i.e., $I(\widetilde{\mathbf{X}}; \mathbf{Y}(t_e)|\mathbf{U}) = I(\mathbf{X}; \mathbf{Y}(t_e)|\mathbf{U})$, and requirements 1 and 3 are preserved under any congruent transformation. If $\mathbf{H}(t_e)$ is singular, the problem can first be reduced in size, since $\mathbf{H}(t_e)$ is diagonal for all t . Thus, from this point on, we will assume $\mathbf{H}(t_e) = \mathbf{I}$.

We provide a constructive proof, and show how one can build a Gaussian input distribution such that all three requirements are fulfilled. We begin by rewriting requirement 1 as a condition on the matrix $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e)$ rather than on the covariance matrix $\mathbf{R}_{\mathbf{X}_G}$. We do so by defining a new matrix, which is the distance of the MMSE matrix $\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t_e)$ from the linear MSE matrix $\mathbf{E}_{\mathbf{X}}^{\text{lin}}(t_e)$. We proceed by showing that, there exists a fraction such that, by defining $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e)$ to be that fraction of the newly defined matrix, we comply also with requirement 2.

As explained above, we begin by rewriting requirement 1 in terms of the matrix $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e)$. Requirement 3 is already a requirement on the matrix $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e)$ and is as follows,

$$\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e) = \mathbf{E}_G(t_e) - \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t_e) \succeq \mathbf{0}. \quad (206)$$

The MMSE for the Gaussian input is:

$$\begin{aligned} \mathbf{E}_G(t_e) &= \mathbf{R}_{\mathbf{X}_G} - \mathbf{R}_{\mathbf{X}_G}(\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1}\mathbf{R}_{\mathbf{X}_G} \\ &= \mathbf{R}_{\mathbf{X}_G} - \mathbf{R}_{\mathbf{X}_G}(\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1}(\mathbf{R}_{\mathbf{X}_G} + \mathbf{I}) + \mathbf{R}_{\mathbf{X}_G}(\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1} \\ &= \mathbf{R}_{\mathbf{X}_G}(\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1} \\ &= (\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})(\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1} - (\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1} \\ &= \mathbf{I} - (\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1}. \end{aligned} \quad (207)$$

From equation (206) $\mathbf{R}_{\mathbf{X}_G}$ complies with the following:

$$(\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1} = \mathbf{I} - \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t_e) - \mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e). \quad (208)$$

Note that the above equation connects $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e)$ with $\mathbf{R}_{\mathbf{X}_G}$. Thus, given a specific substitution of $\mathbf{Q}(\mathbf{X}|\mathbf{U}, \mathbf{R}_{\mathbf{X}_G}, t_e)$ we have a complete definition of the Gaussian input distribution. Similarly, the MMSE assuming an optimal linear estimator of \mathbf{X} (only from $\mathbf{Y}(t_e)$) is given by:

$$\mathbf{E}_{\mathbf{X}}^{\text{lin}} = \mathbf{I} - (\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} \quad (209)$$

and we have that,

$$\mathbf{E}_{\mathbf{X}|\mathbf{U}}(t_e) \preceq \mathbf{E}_{\mathbf{X}}^{\text{lin}}(t_e) \quad \forall t. \quad (210)$$

Thus, we can define:

$$\begin{aligned} \mathbf{C} &\equiv \mathbf{E}_{\mathbf{X}}^{\text{lin}}(t_e) - \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t_e) = \mathbf{I} - (\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} - \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t_e) \succeq \mathbf{0} \\ \mathbf{E}_{\mathbf{X}|\mathbf{U}}(t_e) &= \mathbf{I} - (\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} - \mathbf{C}, \quad \mathbf{C} \succeq \mathbf{0}. \end{aligned} \quad (211)$$

Note that \mathbf{C} is completely defined by the input random vector, \mathbf{X} . Inserting (211) into equation (208) we have:

$$\begin{aligned} (\mathbf{R}_{\mathbf{X}_G} + \mathbf{I})^{-1} &= \mathbf{I} - [\mathbf{I} - (\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} - \mathbf{C}] - \mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e) \\ &= (\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} + \mathbf{C} - \mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e), \quad \mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e) \succeq \mathbf{0}, \quad \mathbf{C} \succeq \mathbf{0}. \end{aligned} \quad (212)$$

We now require the following supporting lemma,

Lemma 19: Assume $\mathbf{X} \in \mathbb{R}^n$ is an arbitrary distributed random vector. For any $t' \in [0, \infty)$ there exists a Gaussian random vector, \mathbf{X}_G , with covariance matrix $\mathbf{R}_{\mathbf{X}_G}$ such that,

- 1) $\mathbf{0} \preceq \mathbf{R}_{\mathbf{X}_G} \preceq \mathbf{R}_{\mathbf{X}}$
- 2) $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t') = \mathbf{0}$
- 3) $I(\mathbf{X}_G; \mathbf{Y}_G(t')) \leq I(\mathbf{X}; \mathbf{Y}(t')|U)$

Proof: See Appendix A10. ■

Note that according to Lemma 19 we have that for $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e) = \mathbf{0}$ there exists a Gaussian random vector, \mathbf{X}_G , which ensures $I(\mathbf{X}_G; \mathbf{Y}_G(t_e)) \leq I(\mathbf{X}; \mathbf{Y}(t_e)|U)$. On the other hand, if $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e) = \mathbf{C}$ we have, according to (212), that $\mathbf{R}_{\mathbf{X}_G} = \mathbf{R}_{\mathbf{X}}$ in which case we have $I(\mathbf{X}_G; \mathbf{Y}_G(t_e)) \geq I(\mathbf{X}; \mathbf{Y}(t_e)|U)$. Moreover, from (212) we can observe that instead of requirement 1 i.e., $\mathbf{R}_{\mathbf{X}_G} \preceq \mathbf{R}_{\mathbf{X}}$, we may simply require $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e) \preceq \mathbf{C}$ (212), thus requirements 1 and 3 can be written as follows,

$$\mathbf{0} \preceq \mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e) \preceq \mathbf{C} \quad (213)$$

where \mathbf{C} is defined in equation (211). The question is whether there exists such a $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}, t_e) \preceq \mathbf{C}$ that will also attain requirement 2 i.e., $I(\mathbf{X}_G; \mathbf{Y}_G(t_e)) = I(\mathbf{X}; \mathbf{Y}(t_e)|U) \equiv \alpha$. From the above mentioned we know that,

$$\frac{1}{2} \log |\mathbf{I} + \mathbf{R}_{\mathbf{X}_G}^1| \leq \alpha \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{R}_{\mathbf{X}_G}^2| \quad (214)$$

where,

$$\mathbf{I} + \mathbf{R}_{\mathbf{X}_G}^1 = ((\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} + \mathbf{C})^{-1} \quad (215)$$

$$\mathbf{I} + \mathbf{R}_{\mathbf{X}_G}^2 = \mathbf{I} + \mathbf{R}_{\mathbf{X}}. \quad (216)$$

Thus, (214) can be rewritten as:

$$\begin{aligned} \frac{1}{2} \log |((\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} + \mathbf{C})^{-1}| &\leq \alpha \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{R}_{\mathbf{X}}| \\ \frac{1}{2} \log \frac{|\mathbf{I}|}{|(\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} + \mathbf{C}|} &\leq \alpha \leq \frac{1}{2} \log \frac{|\mathbf{I}|}{|(\mathbf{I} + \mathbf{R}_{\mathbf{X}})^{-1}|}. \end{aligned} \quad (217)$$

We now need the following result,

Lemma 20: Let's define the function:

$$r(\nu) = \frac{1}{2} \log \frac{|\mathbf{A}|}{|\mathbf{B} + \mathbf{\Delta} \nu|}. \quad (218)$$

For $\mathbf{A} \succ \mathbf{0}$, $\mathbf{B} \succ \mathbf{0}$ and $\mathbf{\Delta} \succeq \mathbf{0}$, the function, $r(\nu)$ is continuous and monotonically decreasing in ν for $0 \leq \nu \leq 1$.

Proof: The proof is similar to the proof of Lemma 10 in [5]. ■

In our case we have:

$$\mathbf{A} = \mathbf{I} \succ \mathbf{0} \quad (219)$$

$$\mathbf{B} = (\mathbf{I} + \mathbf{R}_X)^{-1} \succ \mathbf{0} \quad (220)$$

$$\mathbf{\Delta} = \mathbf{C} \succeq \mathbf{0} \quad (221)$$

and,

$$\frac{1}{2} \log \frac{|\mathbf{A}|}{|\mathbf{B} + \nu \mathbf{\Delta}|} \Big|_{\nu=1} \leq \alpha \leq \frac{1}{2} \log \frac{|\mathbf{A}|}{|\mathbf{B} + \nu \mathbf{\Delta}|} \Big|_{\nu=0}. \quad (222)$$

Thus, according to Lemma 20, there exists a ν^* such that, $r(\nu^*) = \alpha$. That is,

$$\begin{aligned} \alpha &= \frac{1}{2} \log \frac{|\mathbf{A}|}{|\mathbf{B} + \nu^* \mathbf{\Delta}|} = \frac{1}{2} \log \frac{|\mathbf{I}|}{|(\mathbf{R}_X + \mathbf{I})^{-1} + \nu^* \mathbf{C}|} \\ &= \frac{1}{2} \log |\mathbf{I} + \mathbf{R}_{X_G}^*| = \frac{1}{2} \log \frac{|\mathbf{I}|}{|(\mathbf{R}_X + \mathbf{I})^{-1} + \mathbf{C} - \mathbf{Q}(X|U, \mathbf{R}_{X_G}, t_e)^*|} \end{aligned} \quad (223)$$

where the last equality is due to equation (212). That is,

$$\mathbf{Q}(X|U, \mathbf{R}_{X_G}, t_e)^* = (1 - \nu^*) \mathbf{C} \quad (224)$$

and since $0 \leq \nu^* \leq 1$ we have that $\mathbf{0} \preceq \mathbf{Q}(X|U, \mathbf{R}_{X_G}, t_e)^* \preceq \mathbf{C}$, as required. To conclude, we can construct a Gaussian input distribution, complying with all three requirements, as follows,

$$\mathbf{I} + \mathbf{R}_{X_G}^* = ((\mathbf{R}_X + \mathbf{I})^{-1} + \mathbf{C} - \mathbf{Q}(X|U, \mathbf{R}_{X_G}, t_e)^*)^{-1} = ((\mathbf{R}_X + \mathbf{I})^{-1} + \nu^* \mathbf{C})^{-1} \quad (225)$$

where ν^* is derived from the equality in (223). This completes the proof of the lemma. ■

10) Proof of Lemma 19: We first show that there exists a covariance matrix \mathbf{R}_{X_G} such that requirements 1 and 2 are fulfilled. Then, we will show, using contradiction, that requirement 3 is also fulfilled.

First note that, requirement 2 i.e., $\mathbf{Q}(X|U, \mathbf{R}_{X_G}, t') = \mathbf{0}$ completely defines \mathbf{R}_{X_G} :

$$\begin{aligned} \mathbf{I} - (\mathbf{R}_{X_G} + \mathbf{I})^{-1} &= \mathbf{E}_{X|U}(t') \\ (\mathbf{I} - \mathbf{E}_{X|U}(t'))^{-1} - \mathbf{I} &= \mathbf{R}_{X_G} \end{aligned} \quad (226)$$

where we have used the expression in (207), and using the expression in (209) we can show that $\mathbf{I} - \mathbf{E}_{X|U}(t')$ is an invertible matrix since,

$$\mathbf{E}_{X|U}(t') \preceq \mathbf{E}_X^{\text{lin}}(t') = \mathbf{I} - (\mathbf{R}_X + \mathbf{I})^{-1} \prec \mathbf{I}. \quad (227)$$

We now need to check that the first requirement holds:

$$\begin{aligned} \mathbf{R}_{X_G} &= (\mathbf{I} - \mathbf{E}_{X|U}(t'))^{-1} - \mathbf{I} \succeq \mathbf{0} \\ \mathbf{I} - \mathbf{E}_{X|U}(t') &\preceq \mathbf{I} \\ \mathbf{0} &\preceq \mathbf{E}_{X|U}(t') \end{aligned} \quad (228)$$

and,

$$\begin{aligned}
\mathbf{E}_{\mathbf{X}|U}(t') &\preceq \mathbf{E}_{\mathbf{X}}^{\text{lin}}(t') = \mathbf{I} - (\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} \\
(\mathbf{R}_{\mathbf{X}} + \mathbf{I})^{-1} &\preceq \mathbf{I} - \mathbf{E}_{\mathbf{X}|U}(t') \\
\mathbf{R}_{\mathbf{X}} + \mathbf{I} &\succeq (\mathbf{I} - \mathbf{E}_{\mathbf{X}|U}(t'))^{-1} \\
\mathbf{R}_{\mathbf{X}} - (\mathbf{I} - \mathbf{E}_{\mathbf{X}|U}(t'))^{-1} + \mathbf{I} &\succeq \mathbf{0} \\
\mathbf{R}_{\mathbf{X}} &\succeq \mathbf{R}_{\mathbf{X}_G}.
\end{aligned} \tag{229}$$

Thus, we have shown that given an arbitrary input we can find the required $\mathbf{R}_{\mathbf{X}_G}$.

We now want to show that $I(\mathbf{X}_G; \mathbf{Y}_G(t')) \leq I(\mathbf{X}; \mathbf{Y}(t')|U)$. For any Gaussian random vector \mathbf{X}_G^* with covariance $\mathbf{R}_{\mathbf{X}_G}^*$ such that $\mathbf{R}_{\mathbf{X}_G}^* \prec \mathbf{R}_{\mathbf{X}_G}$ we have $\mathbf{E}_G^*(t') \prec \mathbf{E}_G(t')$ and thus, $\mathbf{Q}(\mathbf{X}|U, \mathbf{R}_{\mathbf{X}_G}^*, t') \prec \mathbf{0}$. Using Theorem 8, assuming that we do not have $\mathbf{B}(t) = \mathbf{0}$ for all $0 \leq t \leq t'^4$, and the I-MMSE relationship (47) we have,

$$I(\mathbf{X}_G^*; \mathbf{Y}_G^*(t')) < I(\mathbf{X}; \mathbf{Y}(t')|U). \tag{230}$$

Now let's assume that,

$$I(\mathbf{X}_G; \mathbf{Y}_G(t')) > I(\mathbf{X}; \mathbf{Y}(t')|U). \tag{231}$$

The function $I(\mathbf{X}_G; \mathbf{Y}_G(t'))$ is continuous in the value of its eigenvalues, since,

$$I(\mathbf{X}_G; \mathbf{Y}_G(t')) = \frac{1}{2} \sum_{i=1}^n \log(1 + \lambda_i(\mathbf{R}_{\mathbf{X}_G})). \tag{232}$$

We can construct $\mathbf{R}_{\mathbf{X}_G}^*$ by reducing by ϵ the value of all eigenvalues of $\mathbf{R}_{\mathbf{X}_G}$. According to (231) we can find a small enough ϵ , such that the following inequality still holds:

$$I(\mathbf{X}_G; \mathbf{Y}_G(t')) > I(\mathbf{X}_G^*; \mathbf{Y}(t')) \geq I(\mathbf{X}; \mathbf{Y}(t')|U) \tag{233}$$

but this contradicts (230) and by that proves that,

$$I(\mathbf{X}_G; \mathbf{Y}_G(t')) \leq I(\mathbf{X}; \mathbf{Y}(t')|U). \tag{234}$$

This concludes the proof of the lemma. ■

B. Converse Proof of BC Capacity Under Per-Antenna Constraints for M -Users

We consider the *degraded* parallel Gaussian BC channel:

$$\mathbf{Y}_j[m] = \mathbf{H}_j \mathbf{X}[m] + \mathbf{N}_j[m] \quad j = 1, \dots, M \tag{235}$$

⁴ $\mathbf{B}(t) = \mathbf{0}$ for all $0 \leq t \leq t'$ then all mutual informations equal to zero regardless of the input distribution and the lemma holds trivially.

where $\mathbf{N}_j[m]$, $j = 1, \dots, M$ are standard additive Gaussian noise vectors independent for different time indices m (and can be considered independent of each other), and \mathbf{H}_j , $j = 1, \dots, M$ are diagonal positive semidefinite matrices such that $\mathbf{H}_j \preceq \mathbf{H}_{j+1}$, for all $j = 1, \dots, M-1$. $\mathbf{X} \in \mathbb{R}^n$ is the random input vector and it is assumed independent for different time indices m .

We consider a per-antenna power constraint:

$$[\mathbb{E}\{\mathbf{X}\mathbf{X}^\top\}]_{ii} \leq P_i \quad \forall i, 1 \leq i \leq n. \quad (236)$$

Since we have a *degraded* BC, we can use the single-letter expression given in [21]:

$$R_j \leq I(\mathbf{V}_j; \mathbf{Y}_j | \mathbf{V}_{j-1}) \quad j = 1, \dots, M \quad (237)$$

where \mathbf{V}_j are auxiliary random variables, $\mathbf{V}_M \equiv \mathbf{X}$, $\mathbf{V}_0 \equiv \emptyset$, and the union is over all probability distributions satisfying

$$\mathbf{V}_0 - \dots - \mathbf{V}_{M-1} - \mathbf{V}_M - \mathbf{X} - \mathbf{Y}_M - \mathbf{Y}_{M-1} - \dots - \mathbf{Y}_2 - \mathbf{Y}_1. \quad (238)$$

This is an extension of the proof given for the two user case. We begin by rewriting the single-letter expression (237) as follows:

$$R_j \leq I(\mathbf{X}; \mathbf{Y}_j | \mathbf{V}_{j-1}) - I(\mathbf{X}; \mathbf{Y}_j | \mathbf{V}_j) \quad j = 1, \dots, M \quad (239)$$

and more explicitly:

$$\begin{aligned} R_1 &\leq I(\mathbf{X}; \mathbf{Y}_1) - I(\mathbf{X}; \mathbf{Y}_1 | \mathbf{V}_1) \\ R_2 &\leq I(\mathbf{X}; \mathbf{Y}_2 | \mathbf{V}_1) - I(\mathbf{X}; \mathbf{Y}_2 | \mathbf{V}_2) \\ &\vdots \\ R_{M-2} &\leq I(\mathbf{X}; \mathbf{Y}_{M-2} | \mathbf{V}_{M-3}) - I(\mathbf{X}; \mathbf{Y}_{M-2} | \mathbf{V}_{M-2}) \\ R_{M-1} &\leq I(\mathbf{X}; \mathbf{Y}_{M-1} | \mathbf{V}_{M-2}) - I(\mathbf{X}; \mathbf{Y}_{M-1} | \mathbf{V}_{M-1}) \\ R_M &\leq I(\mathbf{X}; \mathbf{Y}_M | \mathbf{V}_{M-1}). \end{aligned} \quad (240)$$

According to Lemma 4 (and the remark after this lemma) we can construct a diagonal path such that

$$\begin{aligned} \mathbf{H}(t_j) &= \mathbf{H}_j \quad j = 1, \dots, M \\ \mathbf{H}(0) &= \mathbf{0} \end{aligned} \quad (241)$$

with $0 \leq t_1 \leq t_2 \leq \dots \leq t_M$. Now, assume a distribution $P_{\{\mathbf{V}_0 \equiv \emptyset, \mathbf{V}_1, \dots, \mathbf{V}_{M-1}, \mathbf{V}_M \equiv \mathbf{X}\}}$ on the tuple $(\mathbf{V}_0 \equiv \emptyset, \mathbf{V}_1, \dots, \mathbf{V}_{M-1}, \mathbf{V}_M \equiv \mathbf{X})$ with covariance matrix \mathbf{R}_X . We begin by proving the following lemma:

Lemma 21: There exist M independent Gaussian inputs \mathbf{X}_{G_j} , with covariance matrices $\mathbf{\Lambda}_{G_j}$ such that,

$$\begin{aligned} \int_0^{t_j} d_i(\mathbf{X} | \mathbf{V}_j, \mathbf{\Lambda}_{G_j}, \tau) d\tau &= 0 \\ \int_0^{t_{j+1}} d_i(\mathbf{X} | \mathbf{V}_j, \mathbf{\Lambda}_{G_j}, \tau) d\tau &\geq 0, \quad \forall i \quad \forall j = 1, \dots, M-1 \end{aligned} \quad (242)$$

and such that $[\Lambda_{G_j}]_{ii} \leq [\Lambda_{G_{j-1}}]_{ii}$, for $j = 2, \dots, M-1$ and $[\Lambda_{G_1}]_{ii} \leq [\mathbf{R}_X]_{ii}$.

Proof: We will prove the above using induction.

The case of $j = 1$: This is identical to the proof given in Section IV-D.

For a general j : We assume the above holds for j and prove for $j+1$. Due to the Markov relation (238) we have that,

$$d_i(\mathbf{X}|\mathbf{V}_{j+1}, \Lambda_{G_j}, t) = d_i(\mathbf{X}|\mathbf{V}_j \mathbf{V}_{j+1}, \Lambda_{G_j}, t), \quad \forall t \quad (243)$$

and thus,

$$d_i(\mathbf{X}|\mathbf{V}_{j+1}, \Lambda_{G_j}, t) = d_i(\mathbf{X}|\mathbf{V}_j \mathbf{V}_{j+1}, \Lambda_{G_j}, t) \geq d_i(\mathbf{X}|\mathbf{V}_j, \Lambda_{G_j}, t) \quad \forall t \quad (244)$$

where the inequality is, again, due to the Markov relation (238) and the definition of the function $d_i(\mathbf{X}|\mathbf{V}_j, \Lambda_{G_j}, t)$, given in equation (79). This provides us with the following inequality,

$$\int_0^{t_{j+1}} d_i(\mathbf{X}|\mathbf{V}_{j+1}, \Lambda_{G_j}, \tau) d\tau \geq \int_0^{t_{j+1}} d_i(\mathbf{X}|\mathbf{V}_j, \Lambda_{G_j}, \tau) d\tau \geq 0 \quad (245)$$

where the first inequality is due to (244) and the second is due to the induction assumption on j (242). Again, following the same derivation as in the proof in Section IV-D, we know that there exists an independent Gaussian input with covariance $\Lambda_{G_{j+1}}$ such that,

$$\int_0^{t_{j+1}} d_i(\mathbf{X}|\mathbf{V}_{j+1}, \Lambda_{G_{j+1}}, \tau) d\tau = 0 \quad (246)$$

$$\int_0^{t'} d_i(\mathbf{X}|\mathbf{V}_{j+1}, \Lambda_{G_{j+1}}, \tau) d\tau \geq 0 \quad \forall t' > t_{j+1}, \quad \forall i \quad (247)$$

where (247) is true specifically for $t' = t_{j+2}$. Finally, from (246) and (245) and the monotonically increasing property of $d_i(\mathbf{X}|\mathbf{V}_{j+1}, \Lambda_G, \tau)$ in $[\Lambda_G]_{ii}$ (fourth property of Corollary 3), and the fact that it is independent of all other entries in Λ_G , we can conclude that $[\Lambda_{G_{j+1}}]_{ii} \leq [\Lambda_{G_j}]_{ii}$. This concludes the proof of the induction. ■
Now, inserting the above bounds (242) (with the addition of the trivial bound on $I(\mathbf{X}; \mathbf{Y}_1)$, under the per-antenna constraint (236)) into the single-letter expression in (240) we obtain the following outer bound

$$\begin{aligned} R_1 &\leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{P} \mathbf{H}_1^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \Lambda_{G_1} \mathbf{H}_1^T| \\ R_2 &\leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_2 \Lambda_{G_1} \mathbf{H}_2^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_2 \Lambda_{G_2} \mathbf{H}_2^T| \\ &\vdots \\ R_{M-2} &\leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_{M-2} \Lambda_{G_{M-3}} \mathbf{H}_{M-2}^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_{M-2} \Lambda_{G_{M-2}} \mathbf{H}_{M-2}^T| \\ R_{M-1} &\leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_{M-1} \Lambda_{G_{M-2}} \mathbf{H}_{M-1}^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_{M-1} \Lambda_{G_{M-1}} \mathbf{H}_{M-1}^T| \\ R_M &\leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_M \Lambda_{G_{M-1}} \mathbf{H}_M^T|. \end{aligned} \quad (248)$$

where \mathbf{P} is a diagonal matrix with $[\mathbf{P}]_{ii} = P_i$, and Λ_{G_j} are positive semidefinite diagonal matrices such that $\mathbf{0} \preceq \Lambda_{G_M} \preceq \Lambda_{G_{M-1}} \preceq \dots \preceq \Lambda_{G_2} \preceq \Lambda_{G_1} \preceq \mathbf{P}$. The achievability of this outer bound is well-known using superposition coding.

■

C. Converse Proof of BC Capacity Under Covariance Constraints for M-Users

We consider the same setting as in Appendix B, given in (235), but now with a covariance constraint,

$$\mathbf{R}_X \preceq \mathbf{S} \quad (249)$$

where \mathbf{S} is some positive definite matrix.

As in Appendix B, since we have a *degraded* BC, we can use the single-letter expression given explicitly in equation (240), with auxiliary random variables complying with the Markov chain as detailed in (238). Furthermore, we construct a path as was done in equation (241). Now, assume distribution $P_{\{\mathbf{V}_0 \equiv \emptyset, \mathbf{V}_1, \dots, \mathbf{V}_{M-1}, \mathbf{V}_M \equiv \mathbf{X}\}}$ on the tuple $(\mathbf{V}_0 \equiv \emptyset, \mathbf{V}_1, \dots, \mathbf{V}_{M-1}, \mathbf{V}_M \equiv \mathbf{X})$ with covariance matrix \mathbf{R}_X . We begin by proving the following lemma,

Lemma 22: There exist M Gaussian inputs \mathbf{X}_{G_j} , with covariance matrices $\mathbf{R}_{\mathbf{X}_{G_j}}$ such that,

$$\begin{aligned} \int_0^{t_j} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X} | \mathbf{V}_j, \mathbf{R}_{\mathbf{X}_{G_j}}, \tau) \right) d\tau &= 0 \\ \int_0^{t_{j+1}} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X} | \mathbf{V}_j, \mathbf{R}_{\mathbf{X}_{G_j}}, \tau) \right) d\tau &\geq 0, \quad \forall j = 1, \dots, M-1 \end{aligned} \quad (250)$$

and such that $\mathbf{0} \preceq \mathbf{R}_{\mathbf{X}_{G_j}} \preceq \mathbf{R}_{\mathbf{X}_{G_{j-1}}}$, for $j = 2, \dots, M$ and $\mathbf{0} \preceq \mathbf{R}_{\mathbf{X}_{G_1}} \preceq \mathbf{S}$. Furthermore, $\mathbf{Q}(\mathbf{X} | \mathbf{V}_j, \mathbf{R}_{\mathbf{X}_{G_j}}, t_j) \succeq \mathbf{0}$, for all $j = 1, \dots, M$.

Proof: We will prove the above using induction.

The case of $j = 1$: This is identical to the proof given in Section V-E.

For a general j : We assume the above holds for j and prove for $j + 1$. Due to the Markov relation (238) we have that,

$$\mathbf{E}_{\mathbf{X} | \mathbf{V}_{j+1}}(t) = \mathbf{E}_{\mathbf{X} | \mathbf{V}_{j+1}, \mathbf{V}_j}(t) \preceq \mathbf{E}_{\mathbf{X} | \mathbf{V}_j}(t), \quad \forall t \quad (251)$$

from which we can conclude that,

$$\mathbf{Q}(\mathbf{X} | \mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{G_j}}, t) = \mathbf{Q}(\mathbf{X} | \mathbf{V}_j \mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{G_j}}, t), \quad \forall t \quad (252)$$

and thus,

$$\mathbf{Q}(\mathbf{X} | \mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{G_j}}, t) = \mathbf{Q}(\mathbf{X} | \mathbf{V}_j \mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{G_j}}, t) \succeq \mathbf{Q}(\mathbf{X} | \mathbf{V}_j, \mathbf{R}_{\mathbf{X}_{G_j}}, t), \quad \forall t. \quad (253)$$

Since $\mathbf{B}(t)$ is a diagonal positive semidefinite matrix for all t , this leads to,

$$\text{Tr} \left(\mathbf{B}(t) \mathbf{Q}(\mathbf{X} | \mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{G_j}}, t) \right) \geq \text{Tr} \left(\mathbf{B}(t) \mathbf{Q}(\mathbf{X} | \mathbf{V}_j, \mathbf{R}_{\mathbf{X}_{G_j}}, t) \right), \quad \forall t. \quad (254)$$

Now, taking into account the induction assumptions on j , together with (253) and (254) we have,

$$\begin{aligned} \mathbf{Q}(\mathbf{X} | \mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{G_j}}, t_{j+1}) &\succeq \mathbf{Q}(\mathbf{X} | \mathbf{V}_j, \mathbf{R}_{\mathbf{X}_{G_j}}, t_{j+1}) \succeq \mathbf{0} \\ \int_0^{t_{j+1}} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X} | \mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{G_j}}, \tau) \right) d\tau &\geq \int_0^{t_{j+1}} \text{Tr} \left(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X} | \mathbf{V}_j, \mathbf{R}_{\mathbf{X}_{G_j}}, \tau) \right) d\tau \geq 0 \end{aligned} \quad (255)$$

which can also be written as:

$$\begin{aligned} \mathbf{Q}(\mathbf{X}|\mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{Gj}}, t_{j+1}) &\succeq \mathbf{0} \\ I(\mathbf{X}_{Gj}; \mathbf{Y}_{Gj}(t_{j+1})) &\geq I(\mathbf{X}; \mathbf{Y}(t_{j+1})|\mathbf{V}_{j+1}). \end{aligned} \quad (256)$$

These are the two conditions required for Lemma 15, with $\mathbf{X}_G^{ub} \equiv \mathbf{X}_{Gj}$. Thus, according to Lemma 15 there exists a Gaussian random vector with covariance $\mathbf{R}_{\mathbf{X}_{Gj+1}}$ such that,

- 1) $\mathbf{R}_{\mathbf{X}_{Gj+1}} \preceq \mathbf{R}_{\mathbf{X}_{Gj}}$
- 2) $I(\mathbf{X}_{Gj+1}; \mathbf{Y}_{Gj+1}(t_{j+1})) = I(\mathbf{X}; \mathbf{Y}(t_{j+1})|\mathbf{V}_{j+1})$
- 3) $\mathbf{Q}(\mathbf{X}|\mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{Gj+1}}, t_{j+1}) \succeq \mathbf{0}$

Property 2 is equivalent to,

$$\int_0^{t_{j+1}} \text{Tr}(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X}|\mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{Gj+1}}, \tau)) d\tau = 0 \quad (257)$$

and from property 3, Corollary 6, and Theorem 8 we can conclude the following:

$$\begin{aligned} \int_0^{t_{j+2}} \text{Tr}(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X}|\mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{Gj+1}}, \tau)) d\tau &= \int_0^{t_{j+1}} \text{Tr}(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X}|\mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{Gj+1}}, \tau)) d\tau \\ &\quad + \int_{t_{j+1}}^{t_{j+2}} \text{Tr}(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X}|\mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{Gj+1}}, \tau)) d\tau \\ &= 0 + \int_{t_{j+1}}^{t_{j+2}} \text{Tr}(\mathbf{B}(\tau) \mathbf{Q}(\mathbf{X}|\mathbf{V}_{j+1}, \mathbf{R}_{\mathbf{X}_{Gj+1}}, \tau)) d\tau \geq 0. \end{aligned} \quad (258)$$

Together with property 1, this concludes the proof of the induction. \blacksquare

Lemma 22 provides us with Gaussian random vectors \mathbf{X}_{Gj} with covariance matrices $\mathbf{R}_{\mathbf{X}_{Gj}}$ with the following properties:

- 1) $\mathbf{0} \preceq \mathbf{R}_{\mathbf{X}_{Gj}} \preceq \mathbf{R}_{\mathbf{X}_{Gj-1}}$, for $j = 2, \dots, M-1$ and $\mathbf{0} \preceq \mathbf{R}_{\mathbf{X}_{G1}} \preceq \mathbf{S}$.
- 2) $I(\mathbf{X}; \mathbf{Y}_j|\mathbf{V}_j) = \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_j \mathbf{R}_{\mathbf{X}_{Gj}} \mathbf{H}_j^\top|$, for $j = 1, \dots, M-1$.
- 3) $I(\mathbf{X}; \mathbf{Y}_{j+1}|\mathbf{V}_j) \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_{j+1} \mathbf{R}_{\mathbf{X}_{Gj}} \mathbf{H}_{j+1}^\top|$, for $j = 1, \dots, M-1$

Substituting these results into the single-letter expression (240), and defining,

$$\begin{aligned} \mathbf{R}_{G1} &= \mathbf{S} - \mathbf{R}_{\mathbf{X}_{G1}} \\ \mathbf{R}_{Gj} &= \mathbf{R}_{\mathbf{X}_{Gj-1}} - \mathbf{R}_{\mathbf{X}_{Gj}}, \quad \forall j = 2, \dots, M-1 \\ \mathbf{R}_{GM} &= \mathbf{R}_{\mathbf{X}_{GM-1}} \end{aligned} \quad (259)$$

provides the following upper bound,

$$\begin{aligned} R_M &\leq \frac{1}{2} \log |\mathbf{H}_M \mathbf{R}_{GM} (\mathbf{H}_M)^\top + \mathbf{I}| \\ R_j &\leq \frac{1}{2} \log \frac{|\mathbf{H}_j \sum_{l=j}^M \mathbf{R}_{Gl} (\mathbf{H}_j)^\top + \mathbf{I}|}{|\mathbf{H}_j \sum_{l=j+1}^M \mathbf{R}_{Gl} (\mathbf{H}_j)^\top + \mathbf{I}|}, \quad \forall j = 1, \dots, M-1 \end{aligned} \quad (260)$$

where \mathbf{R}_{Gj} are some positive semidefinite matrices such that $\mathbf{0} \preceq \sum_{l=1}^M \mathbf{R}_{Gl} = \mathbf{S}$. This completes the converse proof.

The above upper bounds can be attained simultaneously using a joint Gaussian distribution on the tuple,

$$(\mathbf{V}_0 \equiv \emptyset, \mathbf{V}_1, \dots, \mathbf{V}_{M-1}, \mathbf{V}_M \equiv \mathbf{X}) \quad (261)$$

as follows:

$$\mathbf{V}_j = \mathbf{V}_{j-1} + \mathbf{U}_j \quad (262)$$

where $\mathbf{U}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{G_j})$ for $j = 1, \dots, M$, independent of each other, and where \mathbf{R}_{G_j} are positive semidefinite matrices such that $\sum_{l=1}^M \mathbf{R}_{G_l} \preceq \mathbf{S}$. Thus, we attain the upper bounds for $j = 1, \dots, M-1$ as follows:

$$\begin{aligned} R_j &\leq I(\mathbf{V}_j; \mathbf{Y}_j | \mathbf{V}_{j-1}) \\ &= I(\mathbf{X}; \mathbf{Y}_j | \mathbf{V}_{j-1}) - I(\mathbf{X}; \mathbf{Y}_j | \mathbf{V}_j) \\ &= \frac{1}{2} \log \left| \mathbf{H}_j \sum_{l=j}^M \mathbf{R}_{G_l} (\mathbf{H}_j)^T + \mathbf{I} \right| - \frac{1}{2} \log \left| \mathbf{H}_j \sum_{l=j+1}^M \mathbf{R}_{G_l} (\mathbf{H}_j)^T + \mathbf{I} \right|. \end{aligned} \quad (263)$$

For $j = M$ we obtain the following:

$$\begin{aligned} R_M &\leq I(\mathbf{X}; \mathbf{Y}_M | \mathbf{V}_{M-1}) \\ &= \frac{1}{2} \log \left| \mathbf{H}_M \mathbf{R}_{G_M} (\mathbf{H}_M)^T + \mathbf{I} \right|. \end{aligned} \quad (264)$$

Thus, we have shown that (260) is the capacity region under the covariance constraint. ■

D. Proof of Lemma 17

The proof of this lemma follows the proof of [27, Lem. 4], which is very similar to the well known proof for the capacity region of a *degraded* BC in [1]. The proof of the direct part relies on successive decoding at the stronger user and is practically identical to that found in [1]. We will detail the converse proof only.

Let $\bar{\mathbf{Y}}_{i_j}^j$ denote a sequence of n channel outputs of the i_j 'th realization of user j and let W_j for $j = 1, \dots, M$ denote the message indices. Furthermore, let $\mathbf{Y}_{i_j}^j(l)$ be the l 'th sample of $\bar{\mathbf{Y}}_{i_j}^j$ and $\mathbf{Y}_{i_j}^j(1, \dots, l-1)$ be the set of all samples up to $l-1$ (including). We use similar notation for all other random variables. As the capacity region depends only on the marginals $P_{\mathbf{Y}_{i_j}^j | \mathbf{X}}$ we may assume without loss of generality that indeed the mutual distribution is such that

$$(W_1, \dots, W_M) - \mathbf{X} - \mathbf{Y}_{i_M}^M - \mathbf{Y}_{i_{M-1}}^* - \mathbf{Y}_{i_{M-1}}^{M-1} - \mathbf{Y}_{i_{M-1}(M-2)}^* - \dots - \mathbf{Y}_{i_2}^2 - \mathbf{Y}_{i_2}^* - \mathbf{Y}_{i_1}^1 \quad (265)$$

form a Markov chain for every choice of i_1, i_2, \dots, i_M .

Using Fano's inequality and the fact that W_j are independent messages we can write an upper bound of R_j for any $j = 1, \dots, M$ which holds for every $i_j \in \{1, \dots, K_j\}$:

$$\begin{aligned} R_j &\leq \frac{1}{n} I(W_j; \bar{\mathbf{Y}}_{i_j}^j | W_{j-1}, \dots, W_1) + \delta(n) \\ &= \frac{1}{n} \sum_{l=1}^n I(W_j; \mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)) + \delta(n) \end{aligned} \quad (266)$$

$$\begin{aligned} &= \frac{1}{n} \sum_{l=1}^n \left(h(\mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)) - h(\mathbf{Y}_{i_j}^j(l) | W_j, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)) \right) + \delta(n) \\ &= \frac{1}{n} \sum_{l=1}^n \left(h(\mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)) \right. \\ &\quad \left. - h(\mathbf{Y}_{i_j}^j(l) | W_j, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)) \right) + \delta(n) \end{aligned} \quad (267)$$

$$\begin{aligned} &\leq \frac{1}{n} \sum_{l=1}^n \left(h(\mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)) \right. \\ &\quad \left. - h(\mathbf{Y}_{i_j}^j(l) | W_j, \dots, W_1, \mathbf{Y}_{(j+1)j}^*(1, \dots, l-1), \mathbf{Y}_{i_j}^j(1, \dots, l-1)) \right) + \delta(n) \end{aligned} \quad (268)$$

$$\begin{aligned} &= \frac{1}{n} \sum_{l=1}^n \left(h(\mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)) \right. \\ &\quad \left. - h(\mathbf{Y}_{i_j}^j(l) | W_j, \dots, W_1, \mathbf{Y}_{(j+1)j}^*(1, \dots, l-1)) \right) + \delta(n) \end{aligned} \quad (269)$$

$$\begin{aligned} &= \frac{1}{n} \sum_{l=1}^n \left(h(\mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)) \right. \\ &\quad \left. - h(\mathbf{Y}_{i_j}^j(l) | W_j, \dots, W_1, \mathbf{Y}_{(j+1)j}^*(1, \dots, l-1), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)) \right) + \delta(n) \end{aligned} \quad (270)$$

$$\begin{aligned} &\leq \frac{1}{n} \sum_{l=1}^n \left(h(\mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)) \right. \\ &\quad \left. - h(\mathbf{Y}_{i_j}^j(l) | W_j, \dots, W_1, \mathbf{Y}_{(j+1)j}^*(1, \dots, l-1), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)) \right) + \delta(n) \end{aligned} \quad (271)$$

$$= \frac{1}{n} \sum_{l=1}^n \left(h(\mathbf{Y}_{i_j}^j(l) | \mathbf{V}_{j-1}(l)) - h(\mathbf{Y}_{i_j}^j(l) | \mathbf{V}_{j-1}(l), \mathbf{V}_j(l)) \right) + \delta(n) \quad (272)$$

$$= \frac{1}{n} \sum_{l=1}^n I(\mathbf{V}_j(l); \mathbf{Y}_{i_j}^j(l) | \mathbf{V}_{j-1}(l)) + \delta(n) \quad (273)$$

where $\delta(n) \rightarrow 0$ as $n \rightarrow \infty$. The equality in (266) is due to the chain rule of mutual information. The equality in (267) is due to the Markov chain $(W_1, \dots, W_M) - \mathbf{X} - \mathbf{Y}_{i_j}^j - \mathbf{Y}_{j(j-1)}^*$ and the memoryless nature of the channel, as can be seen in the following identity

$$\begin{aligned} &P \left\{ \mathbf{Y}_{i_j}^j(l) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1) \right\} \\ &= \frac{P \left\{ \mathbf{Y}_{i_j}^j(l), \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1) \right\}}{P \left\{ \mathbf{Y}_{j(j-1)}^*(1, \dots, l-1) | W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1) \right\}} \end{aligned}$$

$$\begin{aligned}
&= \frac{P\left\{\mathbf{Y}_{i_j}^j(l)|W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)\right\} P\left\{\mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)|W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1), \mathbf{Y}_{i_j}^j(l)\right\}}{P\left\{\mathbf{Y}_{j(j-1)}^*(1, \dots, l-1)|W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)\right\}} \\
&= \frac{P\left\{\mathbf{Y}_{i_j}^j(l)|W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)\right\} P\left\{\mathbf{Y}_{j(j+1)}^*(1, \dots, l-1)|\mathbf{Y}_{i_j}^j(1, \dots, l-1)\right\}}{P\left\{\mathbf{Y}_{j(j+1)}^*(1, \dots, l-1)|\mathbf{Y}_{i_j}^j(1, \dots, l-1)\right\}} \\
&= P\left\{\mathbf{Y}_{i_j}^j(l)|W_{j-1}, \dots, W_1, \mathbf{Y}_{i_j}^j(1, \dots, l-1)\right\}.
\end{aligned}$$

The inequality in (268) follows from the fact that conditioning decreases entropy. (269) and (270) follow, again, from the Markov chain $(W_1, \dots, W_M) - \mathbf{X} - \mathbf{Y}_{(j+1)j}^* - \mathbf{Y}_{i_j}^j - \mathbf{Y}_{j(j-1)}^*$ and the memoryless nature of the channel. (271) follows from the fact that conditioning decreases entropy. In (272) we used the following definition of auxiliary random variables:

$$\mathbf{V}_j(l) = \left(W_j, \dots, W_1, \mathbf{Y}_{j(j+1)j}^*(1, \dots, l-1)\right). \quad (274)$$

Next we replace the index l with a random variable I which is uniformly distributed over the integers $1, \dots, n$ and define $\mathbf{V}_j = (\mathbf{V}_j(I), I)$, $\mathbf{X} = \mathbf{X}(I)$, $\mathbf{Y}_{i_j}^j = \mathbf{Y}_{i_j}^j(I)$. As the channel is memoryless, we get

$$R_j \leq I\left(\mathbf{V}_j; \mathbf{Y}_{i_j}^j | \mathbf{V}_{j-1}\right) + \delta(n) \quad (275)$$

for all j and for all i_j . Note that as the channel is memoryless these auxiliary random variables satisfy the Markov chain defined in (160). Moreover, from this definition one can easily see that $\mathbf{V}_0 \equiv \emptyset$ and the largest region will be attained when $\mathbf{V}_M \equiv \mathbf{X}$. Finally, as the above inequalities hold for every $j = 1, \dots, M$ and every $i_j = 1, \dots, K_j$, we complete the proof by taking n to infinity. ■

REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements in Information Theory, Second Edition*. Wiley-Interscience, New York, 2006.
- [2] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, September 2006.
- [3] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Transactions on Information Theory*, vol. 20, pp. 279–280, March 1974.
- [4] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information theoretic problems," in *Proc. IEEE International Symposium on Information Theory (ISIT 2008)*, pp. 1016–1020, Seattle, WA, July 9-14 2006.
- [5] E. Ekrem and S. Ulukus, "The secrecy capacity region of the Gaussian MIMO multi-receive wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 4, April 2011.
- [6] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Transactions on Information Theory*, vol. 57, no. 4, April 2011.
- [7] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.
- [8] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, pp. 101–112, June 1959.
- [9] T. E. Duncan, "On the calculation of mutula information," *SIAM Journal on Applied Mathematics*, vol. 19, no. 1, pp. 215–220, July 1970.

- [10] T. T. Kadota, M. Zakai, and J. Ziv, "Mutual information for the white Gaussian channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 17, no. 4, pp. 368–371, July 1971.
- [11] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1837–1849, May 2008.
- [12] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," submitted to the *IEEE Transactions on Information Theory*, 2011, arXiv:1101.0302v1.
- [13] D. Guo, S. Shamai (Shitz), and S. Verdú, "Additive non-Gaussian noise channels: Mutual information and conditional mean estimation," in *Proc. IEEE International Symposium on Information Theory (ISIT 2005)*, Adelaide, Australia, September 4–9 2005.
- [14] D. P. Palomar and S. Verdú, "Representation of mutual information via input estimates," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 453–470, February 2007.
- [15] S. Verdú, "Mismatch estimation and relative entropy," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3712–3720, August 2010.
- [16] T. Weissman, "The relationship between causal and non-causal mismatched estimation in continuous-time awgn channels," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4256–4273, September 2010.
- [17] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141–154, January 2006.
- [18] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd edition. New York, Wiley, 2007.
- [19] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE International Symposium on Information Theory (ISIT 2006)*, Seattle, WA, July 9–14 2006.
- [20] —, "Estimation in Gaussian noise: Properties of the minimum mean-square error," in *Proc. IEEE International Symposium on Information Theory (ISIT 2008)*, Toronto, ON, Canada, July 6–11 2008.
- [21] T. M. Cover, "Comments on broadcast channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2524–2530, October 1998.
- [22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: University Press, 1985.
- [23] F. Rellich, *Perturbation theory of eigenvalue problems*. New York: Courant Institute of Mathematical Sciences, New York University, 1954. [Online]. Available: <http://www.archive.org/details/perturbationtheo00rell>
- [24] P. Lancaster, "On eigenvalues of matrices dependent on a parameter," *Numerische Mathematik*, vol. 6, no. 1, pp. 377–387, 1964.
- [25] R. Bhatia, *Positive definite matrices*. Princeton University Press, 2007.
- [26] S. M. Kay, *Fundamentals of Statistical Signal processing: Estimation Theory*. Englewood Cliffs, N.J.: Prentice-Hall International.
- [27] H. Weingarten, T. Liu, S. Shamai (Shitz), Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multi-input-multiple-output compound broadcast channel," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5011–5023, November 2009.
- [28] A. W. Marshall and I. Olkin, *Inequalities: theory of majorization and its applications*. Academic Press. New York, 1979.
- [29] R. A. Horn and C. R. Johnson, *Topics on matrix analysis*. Cambridge, UK: University Press, 1991.